

Probability and Hilbert's VI problem.

A. Gandolfi*

email: ag189@nyu.edu - gandolfi@math.unifi.it

Abstract

This work has been prompted by the surprising lack of mathematical coherence in the common usage of some of the fundamental entities in the theory of probability, with an inherent risk of contradiction. While disentangling the intricacies, we realized that the same issue has been raised many times, with only partial solutions, notably by Boole, Hilbert, De Finetti and Renyi, among others. In particular, a restoration of foundational coherence in the usage of probability theory appears to be a missing piece in the solution of Hilbert VI problem.

Here we solve the problem by a new formalization of probability theory based on a minimal collection of axioms with additional context dependent conditions, whose overall consistency is then semantically verified. In Elementary Probability, i.e. probabilities involving boolean combinations of finitely many events, our theory leads to algebraization and, using Tarski Seidenberg reduction, to a proof of decidability of all problems. Inconsistency in Elementary Probability, on the other hand, is equivalent to, suitably redefined, arbitrage or Dutch Book. In the continuous case this leads to nonstandard analysis.

Key words and phrases: probability, discrete probabilities, conditional probabilities, independence, moment problems, finite additivity, existence theorems, Hilbert, Boole, De Finetti, Kolmogorov, axioms, model theory, consistency, elementary probability, Tarski Seidenberg, positivstellensatz, Dutch Books, joint normals.

AMS subject classification: 60C05, 60K35.

*NYU Abu Dhabi and Dipartimento di Matematica e Informatica U. Dini, Università di Firenze, Firenze

Contents

1	Introduction	3
2	Probability	6
2.1	Requirements	6
2.2	Probability environments	8
2.3	Joint perceivability and mutual independence	9
2.4	Arbitrage or Dutch Books	13
3	Elementary Probability	14
3.1	Probabilities involving a finite number of events	15
3.2	Algebraization and decidability of Elementary Probability . .	16
3.3	Relation with semialgebraic geometry	18
4	Elementary Probabilities via Dutch Books	19
4.1	Dutch Books in Elementary Probability	20
4.2	Variations	25
5	Continuous variables	27
5.1	Generalized moments problem and nonstandard analysis . . .	27
5.2	A decidable fragment: poly-moments conditions for joint normals	30
	Appendix A A contradictory set up	37
	Appendix B Analysis of the contradictory setup	39
	Appendix C Example	40
	Appendix D Example of Dutch Book	43
	Appendix E A short cut Dutch Book	45
	Appendix F Paradoxes	46

1. Introduction

There is a long history in the search of a theory of probabilities see e.g. [TF1973, VP94], and Hilbert VI problem [LC2004] calls for its axiomatization; this has been generally interpreted, most of the time by Hilbert himself, as the quest for a collection of few axioms from which the rest of the theory can be derived. Kolmogorov [K1956] proposed one such axiom system which, although disputed by De Finetti and others, has led to a clarification of the foundations, and has become the standard accepted solution.

Yet, there are aspects which, surprisingly, have been mostly overlooked to this day. They concern a lack of mathematical coherence in most of the applications and exercises involving probability theory. Fundamental concepts, among which independence and conditional probabilities, are presented and used in two inconsistent ways: in the theory, they are introduced as definitions, but in the applications they are unfoundedly taken as assumptions. Some confusion about the role of the main probabilistic concepts was recognized by Hilbert, who in 1905 indicates that "at its present state of development, the "axioms" and the "definitions" somewhat overlap with each other" [LC2004, UK2011, DH1905]. The "overlap" has never ceased in applications of probability. In addition, this confusion spawns a potential risk of contradiction, as illustrated, for instance, by the exercise in Appendix A; the consequences of inconsistency could obviously be quite severe in applied contexts, as unwarranted conclusions, for instance about safety, might be drawn from contradictory assumptions.

There have been several other calls and attempts at formalizing probability theory, notably by Bohlmann [GB1901], Keynes [Ke1921], Savage, Popper [Po1938], Renyi [R1955, K84], and theories like Quantum Probability [RS2007, Pit1989b], Free Probability [B03] and Bayesian Probability, raising diverse issues such as the use of sets in Kolmogorov axioms, the significance of countable additivity, and again the role of independence and conditional probabilities; none of these seems to be completely satisfactory. In parallel, the problem of potential contradictions is explicitly mentioned in the works of Boole [B54, Ha1976] and De Finetti [DF1980, DF1974], and later in PSAT [N1986, HJ2000]; but these last researches consider only the linear cases, and hence cannot deal with concepts like independence. To conclude, several paradoxical statements have also been proposed [E2012, Ha2013, Ly2014], often intertwined with the same lack of coherent usage of the basic concepts. The need of a formulation which is able to deal with possible inconsisten-

cies and other issues seems then to still be a missing piece in the solution of Hilbert VI problem.

The possibility of inconsistencies casts a different light into the quest for axiomatization of probability. In fact, a new system of axioms is created every time a new problem is considered, but many problems in the analysis of random phenomena are so immediate that the need of a consistency check seems to be missed during the mathematical formalization. In addition, the foundations of probability theory proposed so far, and, even more, the overall idea that axiomatization is aimed at finding a small collection of far reaching axioms, offer no tool to prevent inconsistencies in applied problems. Indeed, no axiomatization prevented the exercise reported in Appendix A from being considered appropriate.

We seem, therefore, compelled to assign an additional task to the axiomatization of probability; in essence, we need a flexible system which is able to adapt to single problems, indicating both how to prevent inconsistencies and how to preserve the calculative power of probabilistic concepts. This is problem we treat in this paper by proposing a new formalization. We see below that in such formalization concepts like independence and conditional probabilities end up consistently playing a dual role, acting both as assumptions, whose consistency has to be checked, and as definitions, which are the starting points of calculations. As a matter of fact, also additivity is revealed to posses the same type of duality.

At first, the idea was a semantical consistency check: once the hypothesis of a problem have been identified, one has to look for a probability space satisfying all the hypothesis, showing thereby a relative consistency (absolute consistency is essentially ruled out by Gödel's second Incompleteness Theorem [G1931]). This is the procedure suggested by Model Theory, also at the basis of moment problems and PSAT. In Section 3 we develop this direction by introducing an algebraization of Elementary Probability which ultimately leads to show its decidability (a result which seems to fulfill Boole's original claim of having a way of solving "all problems in probability" [Ha1976]). As PSAT is a special case of the algebraic problem we formulate in Elementary Probability, which could be called PPSAT (Polynomial PSAT), this too in NP-complete.

We realized, however, that, in pursuing the above direction, the specific assumptions of each problem and the usual axioms of a probability space end up being treated in the same way (we then name them all "requirements"). This offers the chance to relax the standard axioms, allowing parts of them to

become context dependent. This is done in the paper by starting from very basic probability spaces (related to plausibilities in quantum context), and then introducing the notion of "jointly perceivable" events, a notion whose treatment ends up paralleling that of the standard collective independence. Once the two notions are employed together one can give a coherent foundation to diverse formulations of probability, each one being identified by some requirements which are constantly taken within that formulation, with additional ad hoc requirements in each problem. This is described in Section 2. Section 3 then goes back to Elementary Probability and its algebraization.

In Elementary Probability, we see that if there is no model satisfying all the requirements of a problem, then one can determine a suitably redefined arbitrage mechanisms, or Dutch Book (see Section 4). This generalizes the foundational work of De Finetti, and the Fundamental Theorem of Asset Pricing [DS08]. The construction is based on Stengle's Positivstellensatz [LPR2014], and shows that the assumptions of a problem are consistent if and only if, provided some replicability of the events, it is not possible to extract a sure profit from a believer of those requirements. Outside of Elementary Probability the Dutch Book method encounters some difficulties, as its absence is no longer equivalent to existence of a model, see Section 5.1; this phenomenon is known in other contexts, and seems to require either nonstandard analysis [HL85] or extensions of the concept of arbitrage [DS08].

Our proposed method entails several questions about logic. Following Model Theory [TZ2012, E2006], we need to identify a formal language, a class of structures and correspondence rules; in addition a truth predicate [T44] would be needed to ascertain satisfiability. As there does not seem to be an optimal choice for the language (see, e.g. [V2012]), it appears more reasonable that in our context the language itself is chosen in relation to the requirements, allowing the flexibility of selecting a rich model and proof theory for simple problems, and a more expressive language for more elaborate ones. We do not pursue these considerations further in the present paper.

Summarizing, our proposal, which is to a large extent just a formalization of commonly used procedures, is that the mathematical analysis of probabilities should be reversed: instead of looking for axioms which capture as many situations as possible, one can (quite freely) select a collection of assumptions

(i.e. axioms) for each specific problem and then derive consequences from there, with the sole additional constraint of a preliminary consistency check (via existence of a model).

Notice that, along the way of our formalization, we also forgo the need of having a preliminarily fixed set, a desideratum which has been raised by several authors such as Keynes [Ke1921] or Popper [Po1938], and, in some form, by Tao's ansatz [T2011]. In addition, although we do not present the details here, it is clear that our treatment allows to make a parallel development of various formulations of probability theory, and also of some theories which are close to that of probability, such as Choquet's Capacity or Shafer's Evidence [S1976].

Throughout the paper boldface symbols such as $\mathbf{x} = (x_1, x_2, \dots, x_k)$ indicate vectors whose coordinates are clear from the context; δ_A is the Kronecker delta function of A .

On first reading, it is possible to focus on Elementary Probability by going directly to Appendix A and Section 3.

2. Probability

2.1. Requirements

All requirements will be set on equal footing, but we single out a minimal collection which serves as a basis for the entire theory.

Definition 2.1. *A basic probability space is a triple $(\overline{\Omega}, \overline{\mathbb{A}}, \overline{P})$ where $\overline{\Omega}$ is a set, $\overline{\mathbb{A}}$ is a family of subsets of $\overline{\Omega}$ containing $\overline{\Omega}$ and \emptyset , and \overline{P} is a real valued function on $\overline{\mathbb{A}}$ such that*

- (a) $\overline{P}(\emptyset) = 0$;
- (b) $\overline{P}(\overline{\Omega}) = 1$;
- (c) for every $\overline{A} \subseteq \overline{B} \subseteq \overline{\Omega}$, $\overline{P}(\overline{A}) \leq \overline{P}(\overline{B})$.

Elements \overline{A} of $\overline{\mathbb{A}}$ are called basic events, and \overline{P} is called basic probability.

That these assumptions are not contradictory can be seen with $\overline{P}(\overline{A}) = \delta_{\overline{A}, \overline{\Omega}}$ on any set $\overline{\Omega}$.

Note that basic probabilities appear as "plausibilities" in quantum contexts [F86, FL15]. Note also that we decorate the symbols with a hat as they represent "concrete" structures, i.e. sets.

Standard axiomatizations of Probability Theory identify more axioms, but, as mentioned, we are incorporating any further assumption with the specific case by case ones; we name them all requirements.

Definition 2.2. *A requirement is any statement which can hold for a basic probability space.*

Initial examples of requirements are $P(A) = 1/2$ or there are finitely many events; later, when the theory is developed, requirements take more elaborate forms like constraints on moments, a random variable being a martingale, or a stochastic process satisfying a SPDE. Notice that here we use symbols without bar, to express the fact that requirements are stated before specific basic probability space or random variables are determined.

The interest is in collection of requirements:

Definition 2.3. *A probability pre-environment is a quadruple $((\Omega, \mathbb{A}, \mathcal{P}), \mathcal{R})$, in which Ω is a symbol; \mathcal{P} is a set of symbols containing at least P ; \mathbb{A} is a set of symbols containing at least \emptyset and Ω ; \mathcal{R} is a collection of requirements about the symbols in $(\Omega, \mathbb{A}, \mathcal{P})$.*

Examples of probability pre-environments appear everywhere in the usual development of probability theory, both at abstract levels as assumptions of a theorem, and in problems as collections of hypothesis.

In fact, once a pre-environment is described, consequences can be derived by a deductive calculus. This is the usual modus operandi both for theoretical developments and for applications of probability theory.

Requirements are then stratified, in the sense that once deductions are drawn from some requirements, further concepts can be determined which become the basis of new requirements. For instance, one typical requirement is that \mathbb{A} is a σ -algebra, that there is a function $X : \Omega \rightarrow \mathbb{R}$ measurable with respect to \mathbb{A} ; if some additivity is required for P then one can define integration with respect to P , and then require certain properties for the moments of X . A similar process takes place with independence (see also below). In practice, the introduction of requirements and pre-environments can be seen as a merely terminological clarification of the standard probability theory.

Our new formalization points, however, directly to the fact that all the deductive effort could be groundless if a contradiction is present among the requirements. The next section gives a semantic interpretation which closes the circle of our definitions and insures consistency; later on we discuss what can happen with inconsistency.

2.2. Probability environments

The following definition provides at the same a more meaningful constraint on the type of requirements which can appear in a probability pre-environment, and model theoretical consistency.

Definition 2.4. *A probability environment is a probability pre-environment $((\Omega, \mathbb{A}, \mathcal{P}), \mathcal{R})$ such that there exists a basic probability space $(\overline{\Omega}, \overline{\mathbb{A}}, \overline{P})$ satisfying all the requirements.*

A more precise description of how the constraint are to be satisfied involves an interpretation of the symbols in the pre-environment in terms of elements of the basic probability space. This depends on the type of logic; a schematic description is as follows. First, $\overline{\mathbb{A}}$ contains one element for each member of \mathbb{A} ; next, if, in each of the requirements in \mathcal{R} , $\overline{\Omega}$ replaces Ω , the corresponding members in $\overline{\mathbb{A}}$ replace those in \mathbb{A} , \overline{P} replaces P , and each other symbol in \mathcal{P} is replaced by that of a mathematical entity defined in terms of $(\overline{\Omega}, \overline{\mathbb{A}}, \overline{P})$, then the requirements in \mathcal{R} hold. In such case, the elements of \mathbb{A} are events, P is a probability, the basic probability space is called a (probability) model for the environment. We indicate by Ψ a map which realizes the above correspondence.

Notice that in the above definition events are not sets, the probability is not a function etc.

When consistency of a probability environment is ascertained, then consequences can be consistently derived by inference rules. The model theoretical determination of consistency introduces also the possibility of a semantic sequent calculus. We say that a statement is a **possible consequence** of a probabilistic environment if the statement holds for at least one of the basic probability spaces satisfying the requirements. A statement is a **necessary consequence** if it holds for all the basic probability spaces satisfying the requirements. Any theorem in standard probability theory is a necessary

consequence of any environment in which the hypotheses of the theorem itself are (a necessary consequence of the) requirements.

A particular model theoretical proof method consists in showing by model existence that a certain probability environment $((\Omega, \mathbb{A}, \mathcal{P}), \mathcal{R})$, exists; and then proving that $((\Omega, \mathbb{A}, \mathcal{P}), \mathcal{R}')$ is contradictory, where \mathcal{R}' equals \mathcal{R} plus the negation of a statement s . It follows that s is a necessary consequence of the environment. When applied to Elementary Probability in Section [?] below, this leads to a complete solution.

Example 1. *The Uniform Distribution on n points is a Probability Environment in which the requirements can be taken to be: $\Omega = \{a_1, \dots, a_n\}$; \mathbb{A} contains the $n + 2$ symbols $\{\emptyset, \{a_1\}, \dots, \{a_n\}, \Omega\}$; P is defined on a σ -algebra with all events being jointly perceivable; finally, $P(\{a_k\}) = c$, for each $\{a_k\} \in \mathbb{A}$ and some constant c . To verify that this is indeed a Probability Environment it is enough to take, for instance, $\overline{\Omega} := \{1, \dots, n\}$, $\overline{\mathcal{A}} := \mathcal{P}(\overline{\Omega})$, $\overline{P}(\overline{A}) := |\overline{A}|/n$, and replace each $\{a_k\}$ by $\{k\}$.*

Alternatively: no requirements on Ω ; \mathbb{A} contains (at least) the $n + 2$ symbols $\emptyset, A_1, \dots, A_n, \Omega$; $A_i \cap A_j = \emptyset$; and $P(A_i) = P(A_j)$ for all $i \neq j$, $i, j = 1, \dots, n$. The concrete probability space above is again a model of the environment, with the replacement of A_k by $\{k\}$ (this second formulation satisfies Tao's dogma about extendibility [T2011]).

2.3. Joint perceivability and mutual independence

Specific requirements can be imposed for each different problem, but there are standard ones, such as countable additivity or independence, which set probability theory apart from other theories. The imposition of such requirements is facilitated by suggestive definitions. This has always been the case with independence, which as mentioned plays the role of a requirement in applications, and we now introduce a novel notion for additivity; among other things, it brings about the potential to unify diverse formulations of probability theory.

Definition 2.5. *In a probability environment $((\Omega, \mathbb{A}, \mathcal{P}), \mathcal{R})$, a collection of events $\mathfrak{A} = \{A_i\}_{i \in \mathcal{I}}$, $A_i \in \mathbb{A}$, is **jointly perceivable** (for P) if P is countably additive on $\sigma(\mathfrak{A})$, i.e for every countable subcollection of disjoint events $A_i \in \sigma(\mathfrak{A})$, $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$. The events in \mathfrak{A} are also called jointly perceivable.*

Notice that with this definition, additivity, now renamed joint perceivability, plays both the role of an assumption and that of a definition, as much as independence is now doing. This parallelism is further developed here below.

When the requirements on (Ω, \mathcal{A}, P) are that \mathcal{A} is a σ -algebra of jointly perceivable events, we say that (Ω, \mathcal{A}, P) is a **Kolmogorov probability space**. That these requirements are consistent, and hence define a probability environment, can be verified by taking any $\overline{\Omega}$, $\overline{\mathcal{A}} = \{\overline{\Omega}, \emptyset\}$ and $\overline{P} = \delta_{\overline{\Omega}}$ (the semantic consistency check is essentially verbatim cited from [K1956]).

When the requirements on (Ω, \mathcal{A}, P) are that \mathcal{A} is a algebra and that all finite collections of events are jointly perceivable, we say that (Ω, \mathcal{A}, P) is a **finitely additive probability space**.

In other cases joint perceivability might hold for some but not for all finite collections of events. This is the case in test spaces which appear in Quantum Mechanics [FR72, W09, FL15] as well as in other contexts, e.g. [C10].

The notion of joint perceivability has been phrased in a way that makes it comparable to the slightly adapted usual one of mutual independence.

Definition 2.6. *In a probability environment $((\Omega, \mathbb{A}, \mathcal{P}), \mathcal{R})$, a collection of events $\mathfrak{A} = \{A_i\}_{i \in I}$, $A_i \in \mathbb{A}$ is **mutually independent** if for all disjoint classes \mathfrak{A}_i , $i \in I$, I any set of indices, $\mathfrak{A}_i \subseteq \mathfrak{A}$, P is countably multiplicative on the product $\otimes_{i \in I} \sigma(\mathfrak{A}_i)$, i.e for every countable collection of events $A_i \in \sigma(\mathfrak{A}_i)$, $P(\cap_{i=1}^{\infty} A_i) = \prod_{i=1}^{\infty} P(A_i)$. The events A_i are called *mutually independent*.*

To illustrate the parallelism between the concepts of joint perceivability and mutual independence, we call finitely jointly perceivable a collection \mathfrak{A} of events in which additivity holds for all finite collections of elements of $\sigma(\mathfrak{A})$, and finitely mutually independent a collection \mathfrak{A} for which factorization occurs for all finite products $\otimes_{i=1}^n \sigma(\mathfrak{A}_i)$ of disjoint collections $\mathfrak{A}_i \subset \mathfrak{A}$.

In some cases both joint perceivability and mutual independence are finite: take $\Omega = \mathbb{N}$ and \mathfrak{U} an ultrafilter; then $P(A) = \delta_{\mathfrak{U}}(A)$ is both finitely mutually independent and finitely jointly perceivable on $\mathfrak{A} = \mathcal{P}(\mathbb{N})$, but neither is countable. However, if one is countable and the other finite, then the other is countable too.

Theorem 2.7. *Let $(\Omega, \mathbb{A}, P, \mathcal{R})$ be a probability environment, and $\mathfrak{A} \subseteq \mathbb{A}$. If \mathfrak{A} is jointly perceivable and finitely mutually independent, then it is also (countably) mutually independent.*

If \mathfrak{A} is mutually independent and finitely jointly perceivable, then it is also (countably) jointly perceivable.

Proof. (I) As $(\Omega, \sigma(\mathfrak{A}), P)$ is Kolmogorov, the first statement follows from standard probability theory (see, for instance, [K02] pp. 51, 60).

(II) In the other direction, let $A^0 = A^c$ and $A^1 = A$, and consider

$$\mathcal{F} = \{\cap_{\ell=1}^k A_\ell^{\alpha_\ell}, k \in \mathbb{N}, A_j \in \mathbb{Q}, \alpha_j = 0, 1\} \cup \{\emptyset, \Omega\}.$$

Clearly, $A, B \in \mathcal{F}$ implies $A \cap B \in \mathcal{F}$; $A = \cap_{\ell=1}^k A_\ell^{\bar{\alpha}_\ell} \in \mathcal{F}$ implies $A^c = \cup_{\alpha \neq \bar{\alpha}} \cap_{\ell=1}^k A_\ell^{\alpha_\ell}$; hence, \mathcal{F} is a semialgebra containing \mathfrak{A} and P is finitely additive on \mathcal{F} .

To show that P is countably additive on \mathcal{F} consider $A \in \mathcal{F}$ such that $\cap_{\ell=1}^k A_\ell^{\alpha_\ell} = A = \cup_{j=1}^\infty A(j)$, $A(j) = \cap_{\ell=1}^{k_j} A_\ell^{\alpha_{j,\ell}}(j) \in \mathcal{F}$, $A(j)$ disjoint. We now focus on the countable family

$$\tilde{\mathfrak{A}} = \{A_i^{\alpha_{j,i}}(j), j \in \mathbb{N}, 1 \leq i \leq k_j\} \subseteq \mathfrak{A}.$$

Let's fix an order of the elements of $\tilde{\mathfrak{A}}$ and relabel them B_1, B_2, \dots . Then we consider the map $T : \Omega \rightarrow \{0, 1\}^\mathbb{N}$, such that $T(\omega) = (\delta_{B_1}(\omega), \delta_{B_2}(\omega), \dots)$. T is measurable with respect to the Borel σ -algebra in $\{0, 1\}^\mathbb{N}$, and the σ -algebra $\sigma(\tilde{\mathfrak{A}})$. In fact, for each cylinder $C = C_{i_1, \dots, i_k}^{\beta_1, \dots, \beta_k} = \{\rho \in \{0, 1\}^\mathbb{N} : \rho_{i_m} = \beta_{i_m}\}$ we have $T^{-1}(C) = \cap_{m=1}^k B_{i_m}^{\beta_{i_m}}$. It follows that $\sigma = T(P)$ is a finitely additive probability on $\{0, 1\}^\mathbb{N}$; furthermore, $\sigma(C_{i_1, \dots, i_k}^{\beta_1, \dots, \beta_k}) = P(\cap_{m=1}^k B_{i_m}^{\beta_{i_m}}) = \prod_{m=1}^k P(B_{i_m}^{\beta_{i_m}})$ by independence of the B_j 's under P . Hence, for $H_k = \{0, 1\}$, $\mathcal{A}_k = \mathcal{P}(H_k)$, γ_k the countably additive probability on \mathcal{A}_k such that $\gamma_k(1) = P(B_k)$, $H = H^\infty$, σ is a finitely additive probability on H such that for $D_k \subseteq H_k$

$$\begin{aligned} \sigma(\times_{\ell=1}^\infty D_\ell) &= P(T^{-1}(\times_{\ell=1}^\infty D_\ell)) \\ &= P(\cap_{\ell=1}^\infty T^{-1}(D_\ell)) = \prod_{k=1}^\infty P(T^{-1}(D_k)) = \prod_{k=1}^\infty \gamma_k(D_k) \end{aligned} \quad (1)$$

again by countable independence of P .

These are the conditions used in [D74, PS76], see also [K82], to show that there exists a unique finitely additive probability \tilde{P} , satisfying the further condition (2) below, such that (3) holds for \tilde{P} . As (3) holds for σ , if it satisfies the condition below then $\sigma = \tilde{P} = \otimes_{k=1}^\infty \gamma_k$. Hence, σ is countably

additive on the Borel σ -algebra of H . It follows that P is countably additive on $\sigma(\tilde{\mathfrak{A}})$. The condition to check from [D74, PS76] is that for all clopen subsets D of $\times_{\ell=k+1}^{\infty} H_k$,

$$\sigma(D) = \int_{\times_{\ell=1}^k H_k} \sigma(D(x_1, \dots, x_k)) d \otimes_{\ell=1}^k \gamma_k((x_1, \dots, x_k)) \quad (2)$$

where $D \subseteq H$, $D(x_1, \dots, x_k) = \{z = (z_1, z_2, \dots) \in H : (x_1, \dots, x_k, z_1, z_2, \dots) \in D\}$. In the present case, (2) holds by independence and finite additivity, as

$$\begin{aligned} \sigma(D) &= \sigma(\cup_{(x_1, \dots, x_k)} (D(x_1, \dots, x_k) \cap (\rho_1 = x_1, \dots, \rho_k = x_k))) \\ &= \sum_{(x_1, \dots, x_k)} \sigma((D(x_1, \dots, x_k) \cap (\rho_1 = x_1, \dots, \rho_k = x_k))) \quad (3) \\ &= \sum_{(x_1, \dots, x_k)} \sigma((D(x_1, \dots, x_k))) \gamma_k((\rho_1 = x_1, \dots, \rho_k = x_k)) \\ &= \int_{\times_{\ell=1}^k H_k} \sigma(D(x_1, \dots, x_k)) d \otimes_{\ell=1}^k \gamma_k((x_1, \dots, x_k)) \end{aligned}$$

If \mathfrak{A} is countable then the proof would be finished. For general \mathfrak{A} we observe that the countable additivity of P on $\sigma(\tilde{\mathfrak{A}})$ for each \mathfrak{A} implies that P is countably additive on \mathcal{F} .

(III) Consider now a Stone representation [YH52, S16] in which for a finitely additive probability μ on a measurable space (Ω, \mathcal{A}) , with \mathcal{A} a σ -algebra, there are a compact measurable space $(\hat{\Omega}, \hat{\mathcal{A}})$, and a measurable map $\psi : \Omega \rightarrow \hat{\Omega}$, with $\psi(\Omega)$ dense in $\hat{\Omega}$, such that $\psi(E)$ has a unique extension $\hat{E} \in \hat{\mathcal{A}}$, and there is a unique countably additive probability $\hat{\mu}$ on $(\hat{\Omega}, \hat{\mathcal{A}})$ determined by $\hat{\mu}(\hat{E}) = \psi(\mu)(\psi(E))$ for each $E \in \mathcal{A}$. Notice that the extension is monotone as \hat{E} can be defined as the closure, in a suitable topology, of $\psi(E)$: $E_1 \subseteq E_2$ implies that $\hat{E}_1 \subseteq \hat{E}_2$ as (see [S16] [YH52]). The probability $\hat{\mu}(\hat{\Omega} \setminus \psi(\Omega))$ of the corona $\hat{\Omega} \setminus \psi(\Omega)$ is the deficiency of μ [S16].

(IV) As P is countably additive on a semialgebra \mathcal{F} generating $\sigma(\mathfrak{A})$ from Part (II) above, then it has a unique countably additive extension P^{ca} to $\sigma(\mathfrak{A})$ (by standard extension theorem [K02]).

From Part (III) we have \hat{P} defined, and countably additive, on $\sigma(\hat{\mathfrak{A}})$. Let $\mathcal{L} = \{E \subseteq \Omega : P^{ca}(E) = \hat{P}(\hat{E})\}$. Clearly, $\mathcal{F} \subseteq \mathcal{L}$, as for each $E \in \sigma(\mathfrak{A})$

$P^{ca}(E) = P(E) = \hat{P}(\hat{E})$, and \mathcal{F} is a π -system. Moreover, $A, B \in \mathcal{L}$, $A \subseteq B$ implies

$$\begin{aligned} P^{ca}(B \setminus A) &= P^{ca}(B) - P^{ca}(A) \\ &= \hat{P}(\hat{B}) - \hat{P}(\hat{A}) = \hat{P}(\hat{B} \setminus \hat{A}) \end{aligned}$$

where the last equality holds as $\hat{A} \subseteq \hat{B}$ by the monotonicity of the $\hat{\cdot}$ extension; also for an increasing sequence $A_i \in \mathcal{L}$

$$\begin{aligned} P^{ca}(\cup_{i=1}^{\infty} A_i) &= \lim_i P^{ca}(A_i) \\ &= \lim_i \hat{P}(\hat{A}_i) = \hat{P}(\cup_{i=1}^{\infty} \hat{A}_i) \end{aligned}$$

where again the last equality holds by the monotonicity of the $\hat{\cdot}$ extension. Hence, \mathcal{L} is a λ -system, and the π - λ -theorem implies that $\sigma(\mathfrak{A}) \subseteq \mathcal{L}$. It follows that for all $E \in \sigma(\mathfrak{A})$, $P^{ca}(E) = \hat{P}(\hat{E}) = P(E)$, i.e. $P^{ca} = P$, and P is countably additive on $\sigma(\mathfrak{A})$. \square

This theorem underlines once again the fact that results about independent sequences which are valid in a countably additive setting can be proven in the finitely additive setting as well (see, e.g. [K82]).

2.4. Arbitrage or Dutch Books

If a contradiction is derived, by deductive rules, in a probability pre-environment, then this is inconsistent. This derivation can be eased on some occasions by the method of Arbitrages, or Dutch Books. Informally, a Dutch Book is a rigging strategy in which an individual is lead to believe that a certain game is worth playing, while (s)he is losing some strictly positive amount every time; equivalently, it can be defined as a betting scheme to extract a sure profit from an incoherent agent forced to accept any bet on his betting quotients [V2016]. More formally,

Definition 2.8. *Given a probability pre-environment, a **weak Dutch Book** against the believer of the pre-environment is a an additional random variable V , with expectation operator E , added to the probability pre-environment, with the additional requirements that*

1. *if $X = \mathbb{I}_A$, the indicator function of an event in \mathbb{A} , then $E(\mathbb{I}_A) = P(A)$;*
2. *E is linear on the indicator functions;*

3. if $X \geq Y$ are random variables on which E is defined, then $E(X) \geq E(Y)$;
4. $V \leq 0$;
5. $E(V) > 0$.

A **(strict) Dutch Book** is as above, but with 4. and 5. replaced by 4'. $V \leq -1$ and 5'. $E(V) \geq 0$, respectively. In case a Dutch Book exists we call a believer of the inadmissible requirements an **incorrect evaluator** of probabilities

Example 2. If we require that an event A has $P(A) + P(A^c) = 2$ and A and A^c are jointly perceivable, then let $V = \mathbb{I}_A + \mathbb{I}_{A^c} - 2$. For any basic probability space $(\overline{\Omega}, \overline{\mathbb{A}}, \overline{P})$ and any $\omega \in \overline{\Omega}$, $V(\omega) = -1$, but based on the requirements of the pre-environment $E(V) = P(A) + P(A^c) - 2 = 0$. So V is the a strict Dutch Book.

In a limited form, use of Dutch Books to define probability has been proposed by De Finetti [DF1993, DF1980].

If V is a strict Dutch Book then $V - 1$ is a weak Dutch Book. Moreover, if there is a weak Dutch Book then no basic probability space satisfying the requirements of a pre-environment can exist, as for any random variable V on a basic probability space with $V \leq 0$ it holds that, whatever the definition of expectation, $E(V) \leq 0$ by monotonicity of expected values.

In some cases, such as for finitely many requirements on finitely many events, also the opposite holds, and absence of a Dutch Book guarantees the existence of the environment, see Section 4 below. In general, the situation is more complex: in Section 5.1 below we see that for countably many requirements absence of Dutch Books can be compatible with distributions on hyperreals, while no standard distribution exists.

3. Elementary Probability

In this section we consider the theory of probabilities for finitely many events from the point of view of starting from a collection of assumptions (i.e. requirements for a probability environment) and looking for a model satisfying

them (i.e. checking semantic consistency). After observing that most problems can be expressed in terms of real variables, we define Elementary Probability as the collection of problems involving finitely many algebraic relations, and show its decidability.

3.1. Probabilities involving a finite number of events

A general framework for dealing with finitely many events consists of taking the following requirements for a probability pre-environment (Ω, \mathbb{A}, P) (i.e. fixing some symbols and imposing requirements on them):

1. no requirements on Ω ;
2. \mathbb{A} contains at least $n + 2$ events $\Omega, A_1, \dots, A_n, \emptyset$;
3. all finite collections of events are jointly perceivable under P , i.e. P is fully additive;
4. further requirements on P are determined by a collection of expressions of the form

$$g_r = g_r(P(B_1(A_1, \dots, A_n)), \dots, P(B_{k(r)}(A_1, \dots, A_n))) \triangleleft 0, r \in R, \quad (4)$$

where R is a set of indices of any possible cardinality, the g_r 's are real valued functions, the $B_j(A_1, \dots, A_n)$'s, $j = 1, \dots, k(r)$, are boolean combinations of some of the A_1, \dots, A_n 's, $k(r)$ is an integer, and \triangleleft indicates one of $=, \neq, \geq$ (notice that all other inequalities, including $>$, can be obtained combining relations with the above values of \triangleleft).

Lemma 3.1. *The above family of requirements is semantically consistent, i.e. determines a probability environment, if and only if the following happens.*

For every $j = 1, \dots, k$ let B_j be expressed in disjunctive normal form $B_j = \cup_{\alpha \in \Sigma_j} A^\alpha$ [HM2001] for the appropriate $\Sigma_j \subseteq \Sigma = \{-1, 1\}^n$, $A^{-1} = A^c$, $A^1 = A$ and $A^\alpha = \cap_{i=1}^n A_i^{\alpha_i}$. Consider then the change of variables $x_j = \sum_{\alpha \in \Sigma_j} y_\alpha$, using the 2^n variables $\mathbf{y} = \{y_\alpha\}_{\alpha \in \Sigma}$, one for each of the A^α . Then the family of requirements is admissible if and only if the system of equations and inequalities

$$\begin{cases} g_r(x_1(\mathbf{y}), \dots, x_k(\mathbf{y})) = g_r(\sum_{\alpha \in \Sigma_1} y_\alpha, \dots, \sum_{\alpha \in \Sigma_k} y_\alpha) \triangleleft 0 \\ \sum_{\alpha \in \{-1, 1\}^n} y_\alpha = 1 \\ y_\alpha \geq 0 \end{cases} \quad (5)$$

obtained by the change of variables $x_j = x_j(\mathbf{y})$, together with the additional conditions of normalization and nonnegativity, admits a (real) solution $y = (y_\alpha)_{\alpha \in \Sigma}$.

Proof. Clearly, if there is a concrete probability space $(\bar{\Omega}, \bar{\mathcal{A}}, \bar{P})$ with events $\bar{A}_i \in \bar{\mathcal{A}}$ in one to one correspondence with the A_i 's, satisfying all the requirements, the values $y_\alpha = P(A^\alpha)$ form a set of solutions to the the system $g_r(x_1(\mathbf{y}), \dots, x_k(\mathbf{y})) \leq 0$, as all events are jointly perceivable (i.e. P is fully additive).

Viceversa, if a solution $\mathbf{y} = \{y_\alpha\}_{\alpha \in \Sigma}$ exists, then take $\bar{\Omega} = \{-1, 1\}^n$, for each $\bar{\omega} \in \bar{\Omega}$ let $\bar{P}(\bar{\omega}) = y_{\bar{\omega}}$, $\bar{A}_i = \{\bar{\omega} : \bar{\omega}_i = 1\}$, $\bar{\mathcal{A}}$ equal to the σ -algebra generated by the collection of the \bar{A}_i 's, and, finally, P additive (which implies that all events jointly perceivable. It is easy to verify that $P(\emptyset) = 0, P(\bar{\Omega}) = 1$, P is monotone, and satisfies all the requirements (i.e. there is a basic probability space realizing the environment). The requirements are then admissible are requested. □

Notice that a solution of (4) expressed in terms of the \mathbf{x} variables does not imply consistency of the requirements, as these relations are still missing the requirements about additivity (i.e. joint perceivability) and non negativity of probabilities; only absence of a solution could be used to ascertain inconsistency.

Lemma 3.1 suggests a classification of probability environments for finitely many events in terms of the number of equations and inequalities and the type of functions appearing in them. Some problems, such as maximal entropy, involve uncountably many or non polynomial g_j 's; in most situations, however, the requirements involve only finitely many polynomial equations and inequalities. In addition, in all problems involving macroscopic events P is naturally taken as additive (i.e. all events are jointly perceivable). It is natural to call this class of problems **(Classical) Elementary Probability**.

3.2. Algebraization and decidability of Elementary Probability

In Elementary Probability, Requirement 4. above becomes

- 4'. the requirements on P are determined by a finite collection of expres-

sions of the form

$$\sum_{0 \leq \rho_1, \dots, \rho_k \leq s} a_{\rho_1, \dots, \rho_k}(r) \prod_{j=1}^k (P(B_j(A_1, \dots, A_n)))^{\rho_j} \triangleleft 0, \quad r \in R \quad (6)$$

where the $B_j(A_1, \dots, A_n)$'s are boolean combinations of some of the A_1, \dots, A_n 's, $\triangleleft \in \{\geq, =, \neq\}$, $a_{\rho_1, \dots, \rho_k}(r) \in \mathbb{R}$, $R \subset \mathbb{N}$ is a finite set of integers, the ρ_j 's, s and k are integers.

Corollary 3.2. *The consistency problem for probability environments in Elementary Probability is decidable.*

Proof. If there are only a finite number of equations and inequalities involving polynomial g_r 's, then Lemma 3.1 implies that the admissibility problem is equivalent to the nonemptiness of the semialgebraic set defined by the polynomial relations $g_r = g_r(\mathbf{x}(\mathbf{y})) \triangleleft 0$, $r = 1, \dots, m$, together with $\sum_{\alpha \in \{-1, 1\}^n} y_\alpha = 1$ and $y_\alpha \geq 0$, in the variables y_α 's.

Using Tarski-Seidenberg elimination and Sturm's theorem [BCR1998], the existence of a solution is decidable in a finite number of steps. □

Notice that Tarski-Seidenberg and Sturm's theorems are purely existential results, establishing existence or absence of solutions of polynomial equations and inequalities even in cases in which the solutions cannot be explicitly found.

We now have a procedure to check consistency in Elementary Probability: state the assumptions of a problem or a potential application, write them in algebraic form, check consistency by Tarski-Seidenberg elimination or an alternative algorithm [BPR06], proceed with derivation of (now safely consistent) consequences as usual. As a very simple example, in Appendix B the contradictory problem of Appendix A is formally analyzed by means of algebraization. Another example is in Appendix C.

We can, however, make a further step as probability environments allow model theoretical proofs of necessary consequences of the requirements (i.e. derivation of a consequence if all models satisfy it). Indicating the negation of a relation $g(\mathbf{x}) \triangleleft 0$, with $\mathbf{x} \in \mathbb{R}^d$, by $g(\mathbf{x}) \ntriangleleft 0$, we have:

Corollary 3.3. *If*

$$g_{|R|+1}(P(B_1(A_1, \dots, A_n)), \dots, P(B_k(A_1, \dots, A_n))) \triangleleft 0 \quad (7)$$

is an elementary probability relation, then it is a necessary consequence of the probability environment described by the requirements 1., 2., 3., 4.' if and only if the system (5) augmented by the relation $g_{|R|+1}(x_1(\mathbf{y}), \dots, x_k(\mathbf{y})) \triangleleft 0$ admits no solutions. Such consequentality is then decidable.

Proof. In a probability environment the system (5) has at least one solution. If one such solution is also a solution of the augmented system, then there exists a basic probability space in which the requirements and the negation of (7) hold, hence the statement cannot be a necessary consequence. Viceversa, if no solution of (5) solves the augmented system, then (7) holds in all the basic probability spaces which are probabilistic models of the requirements, hence it is a necessary consequence.

As the existence of solutions is decidable, so is the above deduction rule. \square

This allows to change the last step in the solution of problems in Elementary Probability: express the negation of the potential consequence in algebraic form, use again Tarski Seidenberg elimination or another algorithm to verify that there is no longer a solution. If it is so, then the consequence is proven. See the last part of Appendix C for an example.

Albeit NP complete, the method in Corollary 3.3 solves thus "all problems" in Elementary Probability, at least in principle. The same claim has been made by Boole [B54], without being able to complete his program.

3.3. Relation with semialgebraic geometry

Semialgebraic sets of any degree emerge in discussing satisfiability in Elementary Probability, for instance with mutual independence of many events. On the other hand, it is easy to see that each semialgebraic set included in some nonnegative n -dimensional simplex of the form $\Sigma_k = \{(x_1, \dots, x_k) : x_i \geq 0, \sum_{i=1}^k x_i = 1\}$ can be interpreted as a description of admissibility of requirements for some probability environment with jointly perceivable events (i.e. fully additive probability). It is possible to use, for instance, disjoint events.

Theorem 3.4. *Each semialgebraic set included in some nonnegative simplex of the form Σ_k can be expressed as the set of conditions for satisfiability of a probability environment in classical Elementary Probability.*

Proof. Let $g_r(x_1, \dots, x_k) \triangleleft 0, r = 1, \dots, m + k + 1$, be a system of polynomial relations describing a semialgebraic set included in Σ_k . We can always assume that the last relations are $x_j \geq 0$ for $j = 1, \dots, k$, and $\sum_{j=1}^k x_j = 1$; in addition, we have

$$g_r(x_1, \dots, x_k) = \sum_{\rho_1, \dots, \rho_k: 0 \leq \rho_i \leq s \text{ for } i=1, \dots, k} a_{\rho_1, \dots, \rho_k}(r) \prod_{j=1}^k x_j^{\rho_j} \quad (8)$$

for $r = 1, \dots, m$, where s is the overall maximal degree of any variable in any of the polynomials $g_r, r = 1, \dots, m$, and $a_{\rho_1, \dots, \rho_k}(r)$ are, possibly zero, coefficients.

Next, for the given k , consider the requirements 1., 2., 3. and 4.' for a probability environment in which $n := k$, $B_j = B_j(A_1, \dots, A_k) := A_j \cap \bigcap_{i \neq j} A_i^c$, and the polynomial relations in 4.' are $g_r(P(B_1), \dots, P(B_k)) \triangleleft 0$.

As in Lemma 3.1, consider the variables $x_j := P(B_j), j = 1, \dots, k$ and the variables y_α . For $\alpha(j) = (2\delta_{1=j} - 1, \dots, 2\delta_{k=j} - 1)$ we have $x_j = y_{\alpha(j)}$. The complete system of polynomial relations becomes $g_r(y_{\alpha_1}, \dots, y_{\alpha_k}) \triangleleft 0, r = 1, \dots, m, \sum_{j=1}^k y_{\alpha_j} = 1, \sum_{\alpha} y_{\alpha} = 1$, and $y_{\alpha} \geq 0$ for all α . Combining the last three sets of relations, one gets that necessarily $y_{\alpha} = 0$ for all $\alpha \neq \alpha(j)$ for all j . Hence, only the relations $g_r(y_{\alpha(1)}, \dots, y_{\alpha(k)}) \triangleleft 0, r = 1, \dots, m, \sum_{j=1}^k y_{\alpha(j)} = 1$ and $y_{\alpha(j)} \geq 0$ for $j = 1, \dots, k$ are left, which form a system coinciding with the original one. □

The requirements formulated to reproduce a general semialgebraic set have no real probabilistic content, but depending on the specific case, one can sometimes obtain more meaningful problems.

4. Elementary Probabilities via Dutch Books

In this section we prove that if a finite number of polynomial requirements are stated upon probabilities of boolean combinations of finitely many events, then the requirements determine a probability environment if and only if no Dutch Book can be realized against the believer of such requirements.

Besides its intrinsic interest, a reason for developing such equivalence is that it may be computationally advantageous in certain cases with respect to Tarski Seidenberg elimination or related algorithms [Pa2004, BPT2013, L10]. For the definitions see Section 2.

4.1. Dutch Books in Elementary Probability

Theorem 4.1. *The requirements of an elementary probability pre-environment with n events are not consistent if and only if, assuming that it is possible to realize a, finite but sufficiently large, number of i.i.d., jointly perceivable, copies of the collection of events, it is possible to realize a weak Dutch Book against any incorrect evaluator believing such requirements.*

Some care must be used in interpreting the content of this theorem. When talking about an (incorrect) evaluator of elementary probability we intend that (s)he has determined a phenomenon in which (s)he can identify the various events which enter into the requirements. One of the assumptions in the theorem is that it is possible to find or produce a, finite but sufficiently large, number of phenomena in each of which the evaluator is lead to identify "copies" of the original events, in such a way that the original and all these copies are jointly perceivable and collectively independent.

Proof. Consider requirements of the type 1., 2., 3. and 4'. for a problem in Classical Elementary Probability, involving events $A_{i_1}, i_1 = 1, \dots, n$. Feasibility of the requirements is equivalent, by Corollary 3.2, to nonemptiness of the semialgebraic set defined by the polynomial relations $g_r = g_r((\mathbf{x}(\mathbf{y}))) \triangleleft 0$, $r = 1, \dots, m$, together with $\sum_{\alpha \in \{-1, 1\}^n} y_\alpha = 1$ and $y_\alpha \geq 0$, in the variables y_α 's.

By distinguishing the three possible values of \triangleleft , we can assume that the polynomial relations can be expressed as follows:

$$\begin{cases} f_r(\mathbf{y}) = 0, & r = 1, \dots, m_1 \\ g_r(\mathbf{y}) \geq 0, & r = m_1 + 1, \dots, m_1 + m_2 \\ h_r(\mathbf{y}) \neq 0, & r = m_1 + m_2 + 1, \dots, m_1 + m_2 + m_3 = m'. \end{cases} \quad (9)$$

By the positivstellensatz (see [Kr1964, S1974, BCR1998]), the system has no solution if and only if the following happens. There exists a polynomial F in the ideal generated by the f_r 's in $\mathbb{R}[\mathbf{y}]$, a polynomial G in the cone generated

by the g_r 's in $\mathbb{R}[\mathbf{y}]$ and a polynomial H in the multiplicative monoid generated by the h_r 's in $\mathbb{R}[\mathbf{y}]$ such that

$$F + G + H = 0. \quad (10)$$

More explicitly, there are polynomials $t_r \in \mathbb{R}[\mathbf{y}], r = 1, \dots, m_1$; $s_J \in \mathbb{R}[\mathbf{y}], J \subseteq \{m_1 + 1, \dots, m_1 + m_2\}$ which are sums of squares; and even integers $k_r, r = m_1 + m_2 + 1, \dots, m_1 + m_2 + m_3$, such that

$$v(\mathbf{y}) = \sum_{r=1}^{m_1} t_r f_r + \sum_{J \subseteq \{m_1+1, \dots, m_1+m_2\}} s_J \prod_{r \in J} g_r + \prod_{r=m_1+m_2+1}^{m_1+m_2+m_3} (h_r)^{k_r} = 0 \quad (11)$$

([BCR1998]).

We need to investigate the polynomial (11) as a polynomial in the y_α 's before taking into account that all its coefficients are zero. As such, let ν_α be maximal power of the variable y_α , and consider $\nu = \sum_{\alpha \in \{-1,1\}^n} \nu_\alpha$; next, list the α 's in some fixed order $\alpha_1, \dots, \alpha_{\bar{n}}$; for each $\gamma \in \{1, \dots, \bar{n}\}$ let $\Sigma^{(\alpha_\gamma)}$ be the set of all permutations $\sigma^{(\alpha_\gamma)} = (\sigma_i^{(\alpha_\gamma)}), i = 1, \dots, \nu_{\alpha_\gamma}$ of integers

$$\left\{ \sum_{\gamma'=1}^{\gamma-1} \nu_{\alpha_{\gamma'}} + 1, \dots, \sum_{\gamma'=1}^{\gamma} \nu_{\alpha_{\gamma'}} \right\}. \quad (12)$$

We take ν independent, jointly perceivable copies of the events identified by the incorrect evaluator. We then form a random variable, basically by replacing each occurrence of the variables y_α 's in (11) by the indicator function $\mathbb{I}_{\alpha,(j)}$ that the j -th independent copy, with j to be determined, of the event A^α takes place, and then summing the fully replaced polynomial over all permutations of the indices of the copies. We need to specify how to choose the copy to be used for each replacement. We do this in steps for each selection $\{\sigma^{(\alpha)}\}_{\alpha \in \{-1,1\}^n}$ of a permutation for each α :

1. consider each of the polynomials $t_r f_r, s_J \prod_{r \in J} g_r$ and $\prod_{r=m_1+m_2+1}^{m_1+m_2+m_3} (h_r)^{k_r}$ separately.
2. In each such polynomial u consider one of its factors at a time using the factorization in which they are already expressed (for instance t_r and f_r for those in the ideal);

3. expand out completely each such factor into a sum of monomials; consider each monomial separately and if in such a monomial the variable y_α appears at some power $\overline{m}_\alpha^{(1)}$, then replace it by the product $\prod_{j=1}^{\overline{m}_\alpha^{(1)}} \mathbb{I}_{\alpha,(\sigma_j^{(\alpha)})}$; repeat for all variables in \mathbf{y} . Decorate the symbol of the factor by a tilde to indicate the random variable thus obtained, so that t_r is changed into $\tilde{t}_r = \tilde{t}_r^{\sigma^{(\alpha)}}$, for instance; notice that, although we will drop the dependency, the random variable depends on the fixed permutation $\sigma^{(\alpha)}$.
4. Consider the second factor of u ; repeat the previous step, with this change: if the variable y_α appears at some power $\overline{m}_\alpha^{(2)}$, then replace it by the product $\prod_{j=\overline{m}_\alpha^{(1)}+1}^{\overline{m}_\alpha^{(2)}} \mathbb{I}_{\alpha,(\sigma_j^{(\alpha)})}$.
5. Repeat, always using $\mathbb{I}_{\alpha,(\sigma_j^{(\alpha)})}$ referred to new j 's and hence additional copies, till all factors of u have been changed; notice that the total number of copies of A^α used in the procedure is not greater than ν_α .
6. Consider the next polynomial from the list in point 1., and repeat steps 2.-5. till all polynomials in 1. have been changed.

The above procedure produces a random variable

$$\tilde{u}(\sigma^{(\alpha)}) = \sum_{r=1}^{m_1} \tilde{t}_r \tilde{f}_r + \sum_{J \subseteq \{m_1+1, \dots, m_1+m_2\}} \tilde{s}_J \prod_{r \in J} \tilde{g}_r + \prod_{r=m_1+m_2+1}^{m_1+m_2+m_3} (\tilde{h}_r)^{k_r}.$$

Let then $V = \sum_{\text{all permutations } \{\sigma^{(\alpha)}\}_{\alpha \in \{-1,1\}^n}} \tilde{u}(\sigma^{(\alpha)})$.

We compute the expected value of \tilde{v} according to the incorrect evaluator. Consider one of the polynomials in point 1. after substituting the variables with the indicator functions as above; any two of its factors contain indicator functions which refer to different copies of the space, by (12). Therefore, the incorrect evaluator would consider all factors as independent, and factorize the expected value of the product. Similarly, in each monomial inside each factor, the variables were also substituted with indicator functions which refer to different copies of the space by 3. and the fact that the σ 's are permutations; hence, indicator functions in each monomial are considered independent by the incorrect evaluator. We have thus that he/she would

compute the expected value of the product of the indicator functions which has replaced the variables of a monomial $\prod_{\alpha} y_{\alpha}^{k_{\alpha}}$ as

$$E\left(\prod_{\alpha} \prod_{j \in J_{\alpha}} \mathbb{I}_{\alpha, \sigma_j^{(\alpha)}}\right) = \prod_{\alpha} (P(A^{\alpha}))^{k_{\alpha}}$$

for some set of distinct integers J_{α} of cardinality k_{α} . In addition, all events are jointly perceivable, so that the expectation is linear on the sum of monomials. Hence, the incorrect evaluator would compute $E(\tilde{u}(\sigma^{(\alpha)}))$ as the corresponding polynomial in \mathbf{y} with the y_{α} 's replaced by the $P(A^{\alpha})$'s. This is the value for which he/she thinks that the relations in (9) hold. It follows that, if $\bar{\mathbf{y}}$ indicates the value of \mathbf{y} with the above substitutions, the incorrect evaluator would compute, again by linearity of the expected value due to joint perceivability:

$$\begin{aligned} E(V) &= \sum_{\text{all permutations } \{\sigma^{(\alpha)}\}_{\alpha \in \{0,1\}^n}} E(\tilde{u}(\sigma^{(\alpha)})) \\ &= \sum_{\text{all permutations}} \left(\sum_{r=1}^{m_1} E(\tilde{t}_r) E(\tilde{f}_r) + \sum_{J \subseteq \{m_1+1, \dots, m_1+m_2\}} E(\tilde{s}_J) \prod_{r \in J} E(\tilde{g}_r) \right. \\ &\quad \left. + \prod_{r=m_1+m_2+1}^{m_1+m_2+m_3} (E(\tilde{h}_r))^{k_r} \right) \\ &= \left(\prod_{\alpha} \nu_{(\alpha)}! \right) \left(\sum_{r=1}^{m_1} t_r(\bar{\mathbf{y}}) f_r(\bar{\mathbf{y}}) + \sum_{J \subseteq \{m_1+1, \dots, m_1+m_2\}} s_J(\bar{\mathbf{y}}) \prod_{r \in J} g_r(\bar{\mathbf{y}}) \right. \\ &\quad \left. + \prod_{r=m_1+m_2+1}^{m_1+m_2+m_3} (h_r(\bar{\mathbf{y}}))^{k_r} \right) \\ &\geq \left(\prod_{\alpha} \nu_{(\alpha)}! \right) \prod_{r=m_1+m_2+1}^{m_1+m_2+m_3} (h_r(\bar{\mathbf{y}}))^{k_r} > 0 \end{aligned}$$

from the equalities and inequalities in (9), and the properties of the polynomials s_J and the powers k_r .

We finally evaluate V for each possible realization of events in the collection identified by the incorrect evaluator and in all the copies. First, expand \tilde{v} completely, and then collect all terms corresponding to random variables

which have replaced the same monomial $\prod_{\alpha} y_{\alpha}^{k_{\alpha}}$. For each such monomial, there is a certain number m of terms, with coefficients c_1, \dots, c_m ; we have $\sum_{i=1}^m c_i = 0$ as the corresponding monomial in the expansion of the l.h.s of (11) has zero coefficient. When taken with the indicator functions replacing the variables, the sum is not immediately zero, as the indicator functions refer to different copies. On the other hand, keeping track of the permutation and indicating by $\sigma_j^{\alpha}(i)$ the permutation used in such monomial when the coefficient is c_i , we have that all the random variables related to the same monomial add up to

$$\begin{aligned}
& \sum_{\text{all permutations}} \sum_{i=1}^m c_i \prod_{\alpha} \prod_{j \in J_{\alpha}} \mathbb{I}_{\alpha, \sigma_j^{(\alpha)}(i)} \\
&= \sum_{i=1}^m \sum_{\text{all permutations}} c_i \prod_{\alpha} \prod_{j \in J_{\alpha}} \mathbb{I}_{\alpha, \sigma_j^{(\alpha)}(i)} \\
&= \sum_{i=1}^m c_i (\nu - \sum_{\alpha: J_{\alpha} \neq \emptyset} \nu_{\alpha})! \sum_{\text{permutations: } J_{\alpha} \neq \emptyset} \prod_{\alpha} \prod_{j \in J_{\alpha}} \mathbb{I}_{\alpha, \sigma_j^{(\alpha)}(i)} \\
&= \left((\nu - \sum_{\alpha: J_{\alpha} \neq \emptyset} \nu_{\alpha})! \sum_{\text{permutations: } J_{\alpha} \neq \emptyset} \prod_{\alpha} \prod_{j \in J_{\alpha}} \mathbb{I}_{\alpha, \sigma_j^{(\alpha)}(1)} \right) \sum_{i=1}^m c_i \\
&= 0
\end{aligned}$$

where the penultimate equality derives from the fact that in each product all the terms $\mathbb{I}_{\alpha, \sigma_j^{(\alpha)}(i)}$ refer to the same number of different copies by construction, and hence the sum over all permutations does not depend on i .

Therefore, if the payoff of a game is V , the incorrect evaluator is willing to pay an entry fee to participate, but the game ends up being a draw all the time. This is the weak Dutch Book mentioned in the statement of the theorem. □

Some remarks. The amount of the entry fee cannot be predicted in advance but only determined when the terms in (11) are computed,

It is possible to give bounds on the number of copies of the given events, based on bounds on the number of polynomials used in Stengle's Theorem [LPR2014] and the number of permutations. But such bounds are far from optimal.

The random variable V obtained above, representing the payoff in the Dutch Book, is often not the best possible option, especially because a large number of permutations has been introduced. For an actual determination of a game one can often select the copies more carefully so as to set up a game which is more obviously "advantageous" for the incorrect evaluator. In the Appendix D there is a very simple example, worked out completely including an alternative choice for the V .

One could observe that it is not obvious that the incorrect evaluator is capable of computing $E(V)$ according to his or her own assumptions, and that exploitation of incorrect probability evaluations is a (possibly deplorable) art in itself.

4.2. Variations

Corollary 4.2. *If the inadmissible requirements for a finite number of events contain no strict inequalities then, assuming the possibility of producing mutually independent, jointly perceivable copies of the events, a strict Dutch Book can be realized against any incorrect evaluator believing such requirements.*

Proof. If there are no strict inequalities in the requirements, since normalization and nonnegativity of probabilities correspond also to inequalities which are not strict, there are no strict inequalities in (9). Hence, we can take $h = 1$, and this generates the multiplicative monoid. From (10) we have $F + G = -1$. Following the same construction as in the proof of Theorem 4.1, except for the terms in the multiplicative monoid, one gets

$$V = \sum_{\text{all permutations } \{\sigma^{(\alpha)}\}_{\alpha \in \{-1,1\}^n}} \left(\sum_{r=1}^{m_1} \tilde{t}_r \tilde{f}_r + \sum_{J \subseteq \{m_1+1, \dots, m_1+m_2\}} \tilde{s}_J \prod_{r \in J} \tilde{g}_r \right).$$

The incorrect evaluator now estimates $E(V) = 0$, while for each realization $V = -1$.

With payoff V , then, the incorrect evaluator perceives the game as fair, while losing a constant unit amount. This is the Dutch Book mentioned in the statement of the corollary.

□

Notice that in this case the amount lost is fixed, and as such known in

advance of any calculation about the polynomials determining the Dutch Book.

One way to certify inadmissibility of the requirements is to form a Dutch Book using some of the $g(\mathbf{x}) \triangleleft 0$'s only. This is the case for the example in Appendix A, in which there is a linear subsystem which has no solutions; in such a case a simple linear programming technique leads to a linear combination of the equations certifying the inadmissibility of the requirements, and a Dutch Book can be formed as in the corollary below, with just one copy (see also Appendix E).

In fact, one can produce an inadmissibility certifying Dutch Book by explicitly deriving a contradiction from the requirements and use it to construct the Dutch Book. We say that a polynomial equation or inequality $f(\mathbf{y}) \triangleleft 0$ *implies* the inequality $a \geq b$ between two polynomials $a(\mathbf{y}), b(\mathbf{y})$ if $a - b = t(\mathbf{y})f(\mathbf{y})$ where t is either some polynomial in the variables \mathbf{y} for the case in which \triangleleft is $=$, or t is a sum of squares polynomial for the case in which \triangleleft is \geq . The next corollary formalizes how we deduce contradictions with probabilistic calculations; this is actually a very optimistic description, as we are generally quite limited in deducing contradictions, and are hardly able to use intuition about sum of squares polynomials in probabilistic settings.

Corollary 4.3. *Given the system (9), suppose there are polynomials a_1, \dots, a_n in the variables \mathbf{y} such that*

1. $a_1, a_n \in \mathbb{R}$ (i.e. they do not depend on \mathbf{y});
2. $a_1 < a_n$;
3. for each $k = 1, \dots, n-1$, $a_k(\mathbf{y}) \geq a_{k+1}(\mathbf{y})$ is implied by one of the relations in (9) with no strict inequality.

Then, assuming that it is possible to realize a finite but sufficiently large number of independent, jointly perceivable copies of the collection of events, one can form a Dutch Book as follows. Suppose that at step k the relation $f_{i_k} \triangleleft 0$ implies $a_k \geq a_{k+1}$, and let t_k be such that $a_{k+1}(\mathbf{y}) - a_k(\mathbf{y}) = t_k(\mathbf{y})f_{i_k}(\mathbf{y})$. Then $v(\mathbf{y}) = \frac{1}{a_n - a_1} \sum_{k=1}^{n-1} t_k f_{i_k} = -1$; and the random variable V obtained as in the proof of Theorem 4.1 is the payoff a Dutch Book.

Proof. Consider the following procedure. Start from a_1 ; if $f_{i_1} \triangleleft 0$ implies $a_1 \geq a_2$, then let t_1 be the polynomial such that $a_1 - a_2 = t_1 f_{i_1}$; then $a_1 = t_1 f_{i_1} + a_2$;

continuing one gets $a_1 = t_1 f_{i_1} + t_2 f_{i_2} + a_3 = \dots = \sum_{k=1}^{n-1} t_k f_{i_k} + a_n$. Thus $\sum_{k=1}^{n-1} t_k f_{i_k} = a_1 - a_n < 0$ and $v(\mathbf{y}) = -1$. By replacing the \mathbf{y} variables with indicator functions as in the proof of Theorem 4.1 one gets the payoff random variable of a Dutch Book, since $t_k f_{i_k} \geq 0$ in all cases, and $a_n > a_1$. \square

An example is in Appendix E.

5. Continuous variables

Probabilities in the continuous case are characterized by the requirements that Ω is some Borel subset of \mathbb{R} or \mathbb{R}^n , and $\mathbb{A} = \mathcal{B}_\Omega$, the Borel σ -algebra of \mathbb{R} or \mathbb{R}^n restricted to Ω .

5.1. Generalized moments problem and nonstandard analysis

In the classic problem of moments one assigns a Borel subset Ω of \mathbb{R} or \mathbb{R}^n and potential moments, and then looks for existence of random variables (which are described by the requirements of being measurable real valued functions) defined on Ω and satisfying the prescribed moments with respect to the Lebesgue measure [ST1943, L10]. As the unknown is the distribution of the random variable, this is a linear problem.

The solution is generally expressed in terms of Dutch Books, albeit apparently not using this explicit terminology. In fact, the Riesz-Haviland theorem, which can be used to identify the main conditions for existence of solutions to moment problems, states that given values $m_{\mathbf{k}}$, $\mathbf{k} \in \mathbb{N}^n$ and a closed set $\Omega \subseteq \mathbb{R}^n$, there is a probability μ concentrated on Ω such that $\int_\Omega \mathbf{x}^{\mathbf{k}} \mu d\mathbf{x} = m_{\mathbf{k}}$ if and only if the following happens: for a polynomial $p(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbb{N}^n} a_{\mathbf{k}} \prod_{i=1}^n x_i^{k_i}$ in the variables $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{m} = \{m_{\mathbf{k}}\}_{\mathbf{k} \in \mathbb{N}^n}$, we indicate $p(\mathbf{m}) = \sum_{\mathbf{k} \in \mathbb{N}^n} a_{\mathbf{k}} m_{\mathbf{k}}$; then $p(\mathbf{m}) \geq 0$ for every polynomial p such that $p(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \Omega$. Suppose now that an incorrect evaluator of probabilities assigns moments $m_{\mathbf{k}}$ to some random variables \mathbf{X} taking values in $\Omega \subseteq \mathbb{R}^n$, but there is no probability environment for these requirements. Then let $\bar{p}(\mathbf{x})$ be the polynomial, nonnegative on Ω , for which the above condition fails, i.e. $\bar{p}(\mathbf{m}) < 0$. Then, $V(\mathbf{X}) = -\bar{p}(\mathbf{X})$ is a weak Dutch Book against the incorrect evaluator; in fact, s(he) evaluates $E(V) = E(-\bar{p}(\mathbf{X})) = -\bar{p}(\mathbf{m}) > 0$, while in fact, for every possible value \mathbf{x} of the random variable \mathbf{X} , $V = -\bar{p}(\mathbf{x}) \leq 0$.

This type of results can be easily generalized to situations in which, instead of giving directly the (potential) moments of a distribution, polynomial relations between such moments are assigned, provided that the relations involve compact sets.

Theorem 5.1. *Consider a countable collection of polynomials $\mathbf{p} = p_i(\mathbf{m})$, $i \in \mathbb{N}$, in the variables $\mathbf{m} = \{m_{\mathbf{k}}\}_{\mathbf{k} \in \mathbb{N}^n}$, and requirements on a probability μ that it is concentrated on a closed subset $\Omega \subseteq \mathbb{R}^n$ and its moments $m_{\mathbf{k}} = \int_{\Omega} \mathbf{x}^{\mathbf{k}} \mu d\mathbf{x}$ satisfy $p_i(\mathbf{m}) \triangleleft_i 0$ for all i . If the \triangleleft_i 's contain no strict inequalities and there exists constants $M_{\mathbf{k}}, \mathbf{k} \in \mathbb{N}^n$, such that if \mathbf{m} is a solution of all the $p_i(\mathbf{m}) \triangleleft_i 0$'s then*

$$m_{\mathbf{k}} \leq M_{\mathbf{k}} \text{ for each } \mathbf{k} \in \mathbb{N}^n, \quad (13)$$

then there is a probability μ satisfying the requirements if and only if there is no Dutch Book against the believer of the above relations.

In the classic moment problem all \triangleleft_i 's are equalities, hence the relations contain no strict inequalities and the $m_{\mathbf{k}}$'s are bounded; the condition is also fulfilled if, for instance, all equations are of the form $a_{\mathbf{k}} \geq (m_{\mathbf{k}} - b_{\mathbf{k}})^2$, but not if there are some $a_{\mathbf{k}} < m_{\mathbf{k}}^2$.

Proof. Let $C^{(s)}$ be the set of $m_{\mathbf{k}}^{(s)}$ which are solutions of the relations $p_i(\mathbf{m}) \triangleleft_i 0$, $i = 1, \dots, s$, and also satisfy (13); as there are no strict inequalities, the $C^{(s)}$'s form a decreasing sequence of compact sets in \mathbb{R}^{∞} with product topology; and, as (13) holds for all solutions of all relations $p_i(\mathbf{m}) \triangleleft_i 0$, each such solution belongs to $C^{(s)} = \bigcap_{i=1}^s C^{(i)}$ for each s , and hence to $\bigcap_{s \in \mathbb{N}} C^{(s)}$. The absence of any probability satisfying the requirements might occur for two reasons: either $\bigcap_{s \in \mathbb{N}} C^{(s)} = \emptyset$, or for each $\mathbf{m} \in \bigcap_{s \in \mathbb{N}} C^{(s)}$ there is no solution to the moment problem with $\mathbf{m} = \{m_{\mathbf{k}}\}_{\mathbf{k} \in \mathbb{N}}$ as values for the moments.

If $\bigcap_{s \in \mathbb{N}} C^{(s)} = \emptyset$ then there is an s such that $C^{(s)}$ is empty by completeness of \mathbb{R} . Which means that necessarily there is no solution to the polynomial relations $p_i(\mathbf{x}) \triangleleft_i 0$, $i = 1, \dots, s$. By the Positivstellensatz, there is a polynomial $v(\mathbf{x}) = -1$ such that the relations $p_i(\mathbf{x}) \triangleleft_i 0$, $i = 1, \dots, s$ imply $v \geq 0$. We can now perform a construction analogous to the one in the proof of Theorem 4.1, using enough mutually independent, jointly perceivable copies of the random variables on Ω which have been identified by the incorrect evaluator; as the expectation factorizes over the product of random variables depending on mutually independent, jointly perceivable probability spaces, it is easily seen that we get a strict Dutch Book V .

If $\cap_{s \in \mathbb{N}} C^{(s)} \neq \emptyset$, then for each $\tilde{\mathbf{m}} \in \cap_{s \in \mathbb{N}} C^{(s)}$ there is no solution to the moment problem with the $\tilde{m}_{\mathbf{k}}$'s as values for the moments; hence for each such $\tilde{\mathbf{m}}$ there is a polynomial $p^{(\tilde{\mathbf{m}})}$, nonnegative on Ω , for which $p^{(\tilde{\mathbf{m}})}(\tilde{\mathbf{m}}) < 0$. Each such polynomial determines an open set $B^{(\tilde{\mathbf{m}})} = \{\mathbf{m} : p^{(\tilde{\mathbf{m}})}(\mathbf{m}) < 0\}$; and

$$\{B^{(\tilde{\mathbf{m}})}\}_{\tilde{\mathbf{m}} \in \cap_{s \in \mathbb{N}} C^{(s)}}$$

is an open cover of the compact set $\cap_{s \in \mathbb{N}} C^{(s)}$, from which we can extract a finite subcover $\{B^{(\tilde{\mathbf{m}}^{(j)})}\}_{j=1, \dots, r}$. The function $v(\mathbf{x}) = \min_{j=1, \dots, r} p^{(\tilde{\mathbf{m}}^{(j)})}(\mathbf{x})$ is nonnegative on Ω , and yet $v(\mathbf{m}) < 0$ in each solution $\mathbf{m} \in \cap_{s \in \mathbb{N}} C^{(s)}$. Consider payoffs $V = -v(\mathbf{X})$: then $V \leq 0$ for every realization \mathbf{x} of \mathbf{X} ; on the other hand, whatever \mathbf{m} the incorrect evaluator deems the moments to be, it will be $p^{(\tilde{\mathbf{m}})}(\mathbf{m}) < 0$ for some $\tilde{\mathbf{m}}$, so

$$\begin{aligned} E(V(X)) &= E(\max_{j=1, \dots, r} -p^{(\tilde{\mathbf{m}}^{(j)})}(\mathbf{X})) \\ &\geq E(-p^{(\tilde{\mathbf{m}})}(\mathbf{X})) = -p^{(\tilde{\mathbf{m}})}(\mathbf{m}) > 0. \end{aligned}$$

Hence, V is a Dutch Book. □

The situation changes if there are strict inequalities in the $p_i(\mathbf{m}) \triangleleft_i 0$'s or the set of solutions is unbounded, as the set of possible moment values is no longer compact. In such case, absence of a Dutch Book is compatible with absence of a probability distribution satisfying the given constraints. In fact, there might be a Loeb distribution on nonstandard reals [T2012] which satisfies all the requirements.

Example 3. Let $n = 1$; p_0 be $m_1 > 0$; p_{2r} be $m_1 \leq 1/r$; and p_{2r+1} be $(m_1)^r - m_r = 0$, for $r \in \mathbb{N}$. Clearly, there is no standard solution, but the atomic Loeb distribution concentrated on the hyperreal $(1, 1/2, \dots, 1/n, \dots)/\mathcal{U} \in {}^*\mathbb{R}$, where \mathcal{U} is a fixed ultrafilter, satisfies all the requirements.

It is likely to be the case that in the continuous case absence of a Dutch Book is equivalent to the existence of a distribution on nonstandard numbers which satisfies all the requirements.

On the other hand, if one wants to find a structure whose absence guarantees existence of a standard solution to the moment problem in general form, i.e. with possible strict inequalities, one would need to develop a modified

version of Dutch Books: this situation, limited to the existence of a martingale measure in absence of a modified version of arbitrage, has been solved in [DS08].

5.2. A decidable fragment: poly-moments conditions for joint normals

Consistency of some collections of requirements is more manageable; in particular, satisfiability may become decidable in some other classes of problems besides Elementary Probability. We briefly present one example below; it is again based on semialgebraic geometry.

Consider the following requirements. $\Omega = \mathbb{R}$, $\mathbb{A} = \mathcal{B}_{\mathbb{R}}$, and $P = e^{-(\sum_{j=1}^n x_j^2)/2} \lambda$, where λ is the Lebesgue measure; moreover, there are n random variables X_1, \dots, X_n satisfying:

$$X_i = \sum_{j=1}^n a_{i,j} Z_j + b_j \text{ for some } a_{i,j}, b_j \in \mathbb{R} \quad (14)$$

where Z_j are such that their joint distribution has density $dP/d\lambda$ (in short, the Z_j 's are i.i.d. $N(0, 1)$ and the X_i 's have joint normal distributions); finally, the X_i 's satisfy

$$\sum_{0 \leq \rho_i \leq s, \text{ for } i=1, \dots, n} \rho_{k_1, r, \dots, k_n, r}(r) E\left(\prod_{i=1}^n X_i^{k_{i,r}}\right) \triangleleft 0 \quad (15)$$

for $r = 1, \dots, R$, where the $\rho_{k_1, r, \dots, k_n, r}(r)$'s are given real constants, and the $k_{i,r}$'s and s are given integers. We call these probability environments poly-moment conditions for joint normals. We have

Theorem 5.2. *For given ρ_r 's and $k_{i,r}$'s, it is decidable whether the requirements for joint normal distributions determine a probability environment or not.*

Proof. Substitute in (15) the expressions of X_i 's from (14) and expand. The Z_j 's are independent, and their moments are known: $E(Z^k) = (k-1)!!$ for k even, and 0 otherwise. Therefore, we get R polynomial relations in the variables $a_{i,j}$ and b_j , for which the existence of a real solution is decidable as described before. Once we have one selection of values for the $a_{i,j}$'s and the b_j 's, (14) gives the required concrete random variables. \square

Notice that, to the opposite of Elementary Probability, it is not clear under which conditions a system of polynomial relations is obtained from the feasibility test of poly-moment conditions for joint normals.

References

- [AT2006] A. H-S. Ang, W. H. Tang (2006), Probability concepts in engineering, Wiley ed., 2nd edition
- [A1989] V. I. Arnold, Mathematical methods of classical mechanics, Second Edition, Graduate. Texts in Mathematics 60, Springer-Verlag, New York-Berlin, 1989.
- [AP2009] D. Atkinson, J. Peijnenburg (2009) Justification by an Infinity of Conditional Probabilities. Notre Dame Journal of Formal Logic 50,18393.
- [BPR06] Saugata Basu, Richard Pollack, and Marie-Francoise Roy. 2006. Algorithms in Real Algebraic Geometry (Algorithms and Computation in Mathematics). Springer-Verlag New York, Inc., Secaucus, NJ, USA.
- [B03] P. Biane, Free probability for probabilists, Quantum Probability Communications, Volume XI, 55-72, World Scientific (2003)
- [B2010] N.H. Bingham (2010), Finite additivity versus countable additivity: De Finetti and Savage. Electronic J. History of Probability and Statistics 6.1
- [BCR1998] Bochnak, Jacek; Coste, Michel; Roy, Marie-Françoise. Real Algebraic Geometry. Translated from the 1987 French original. Revised by the authors. Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], 36. Springer-Verlag, Berlin, 1998.
- [GB1901] G. Bohlmann Lebensversicherungsmathematik, Encyklopadie der mathematischen Wissenschaften, Bd I, Teil 2, Artikel I D 4 b (1901), 852-917
- [BPT2013] G. Blekherman, P. A. Parrilo, R. R. Thomas (eds.), Semidefinite Optimization and Convex Algebraic Geometry, MOS-SIAM (2013).

- [B54] Boole, George (1854). An Investigation of the Laws of Thought on Which are Founded the Mathematical Theories of Logic and Probabilities. Macmillan. Reprinted with corrections, Dover Publications, New York, NY, 1958.
- [CG2016] Catalano, C., Gandolfi, A. (2016): Partially independent random variables. Preprint.
- [C10] Chichilnisky Graciela (2010) The Foundations of Statistics with Black Swans. Mathematical Social Sciences, 59, 184-192.
- [LC2004] L. Corry, David Hilbert and the Axiomatization of Physics (1898-1918), Kluwer Academic Publishers, Dordrecht (2004)
- [DF1949] de Finetti, Bruno (1949): Sull'impostazione assiomatica del calcolo delle probabilit. Annali Triestini **19** 29-81.
- [DF1993] de Finetti, B. (1993). Sul significato soggettivo della probabilità. (in Italian), Fundamenta Mathematicae, 17, 298-329, 1931. Translated into English as "On the Subjective Meaning of Probability (pp. 291 - 321)". In P. Monari and D. Cocchi (Eds.), Probabilità e Induzione. Bologna: Clueb.
- [DF1980] de Finetti, B. (1980). La prévision: ses lois logiques, ses sources subjectives. Annales de l'Institut H. Poincaré, 7 (pp. 1-68), 1937. Translated into English by Henry E. Kyburg Jr., as Foresight: Its Logical Laws, its Subjective Sources. In H. E. Kyburg Jr. and H. E. Smokler (Eds.), Studies in subjective probability. Wiley: New York, 1964. Second edition published by Krieger, New York, pp. 53-118.
- [DF1974] de Finetti, B. (1974). Theory of Probability, Vol.1. Chichester: Wiley.
- [DS08] Delbaen F. and W. Schachermayer, ?The mathematics of Arbitrage? Springer Finance, 2008.
- [DS1965] Dubins, L. E.; Savage, L. J. (1965). How to gamble if you must: inequalities for stochastic processes. McGraw-Hill series in probability and statistics. McGraw-Hill.

- [D74] L. E. Dubins, On Lebesgue like extensions of finitely additive measures, Ann. Probab. 2 (1974), 456-463.
- [E2012] William Eckhardt (2012): Paradoxes in Probability Theory. Springer.
- [TF1973] Fine, T (1973): Theories of Probability: An Examination of Foundations, New York, Academic Press.
- [F86] Peter C. Fishburn (1986): The Axioms of Subjective Probability. Statistical Science 1(3), pp. 335-358.
- [FL15] Tobias Fritz, Matthew Leifer: Plausibility measures on test spaces arXiv:1505.01151 [quant-ph] (2015).
- [FR72] D. J. Foulis , C. H. Randall (1972): Operational Statistics. I. Basic Concepts. Journal of Mathematical Physics 13(11), pp. 1667-1675,
- [G1931] Kurt Gödel (1931), "Über formal unentscheidbare Stze der Principia Mathematica und verwandter Systeme, I." Monatshefte für Mathematik und Physik 38, 173-198.
- [Ha1965] Hailperin, T. (1965) Best possible inequalities for the probability of a logical function of events. American mathematical monthly, vol. 72, pp. 343-359.
- [Ha1976] Hailperin, T. (1976/1986). Boole's Logic and Probability. North Holland.
- [HJ2000] P. Hansen and B. Jaumard (2000): Probabilistic satisfiability. In Handbook of Defeasible Reasoning and Uncertainty Management Systems, vol.5. Springer.
- [HL85] Hurd, A.E. and Loeb, P.A.: An introduction to nonstandard real analysis, London, Academic Press, 1985
- [HM2001] Hazewinkel, Michiel, ed. (2001), "Disjunctive normal form", Encyclopedia of Mathematics, Springer,
- [DH1905] David Hilbert, Logische Prinzipien des Mathematischen Denkens, Manuskript von Hilberts Vorlesung von 1905, Bibliothek des Math.Inst. Univ. Gottingen. aufgezeichnet von E.Hellinger, 1905.

- [E2006] Richard L. Epstein Classical Mathematical Logic: The Semantic Foundations of Logic, Princeton University Press.
- [Ha2013] Olle Häggström, (2013): Book Review: Paradoxes in Probability Theory, Notices of the AMS, 329-331.
- [Hu1985] Humphreys, P. (1985). Why Propensities Cannot Be Probabilities. Philosophical Review, 94, 557-70
- [K02] O. Kallenberg, Foundations of Modern Probability, Second Edition, Springer, 2002.
- [K84] Kamiski, Andrzej. "On the Rnyi theory of conditional probabilities." Studia Mathematica 79.2 (1984): 151-191
- [K82] R. Karandikar, A general principle for limit theorems in finitely additive probability. Trans. Amer. Math. Soc. Vol. 273, No. 2 (1982).
- [Ke1921] Keynes, John Maynard (1921), Treatise on Probability, London: Macmillan and Co.
- [K1956] Kolmogorov, A. N. (1956): Foundations of the Theory of Probability. Second edition, Chelsea Publ., New York.
- [Kr1964] Krivine, J.L. (1964). "Anneaux préordonnés ". Journal d'analyse mathématique 12, 307-326
- [UK2011] U. Krengel, ON THE CONTRIBUTIONS OF GEORG BOHLMANN TO PROBABILITY THEORY, *Electronic Journal for History of Probability and Statistics* **7 No. 1**, 1-13 (2011).
- [L10] J.-B. Lasserre (2010): Moments, Positive Polynomials and Their Applications, Imperial College Press, London.
- [LPR2014] H. Lombardi, D. Perrucci, M.-F. Roy (2014): An elementary recursive bound for effective Positivstellensatz and Hilbert 17-th problem, Preprint, <http://arxiv.org/abs/1404.2338>
- [Ly2010] Lyon, Aidan (2010): Philosophy of Probabilty. In Fritz Allhoff (ed.), Philosophies of the Sciences: A Guide. Wiley-Blackwell.

- [Ly2014] A. Lyon (2014) : From Kolmogorov, to Popper, to Renyi: There's No Escaping Humphreys' Paradox (When Generalised) in *Chance and Temporal Asymmetry*, Alastair Wilson (ed.), Oxford University Press.
- [N1986] Nils Nilsson (1986) Probabilistic logic. *Artificial Intelligence*, 28(1), 71-87.
- [Pa2004] Parrilo, Pablo A. "Sum of squares programs and polynomial inequalities." *SIAG/OPT Views-and-News: A Forum for the SIAM Activity Group on Optimization*. Vol. 15. No. 2. 2004.
- [Pit1989a] I. Pitowsky I (1989) From George Boole to John Bell: the origin of Bell's inequality. In: Kafatos M (ed.) *Bell's theorem, quantum theory and the conceptions of the universe*. Kluwer, Dordrecht, pp 37-49
- [Pit1989b] Pitowsky, I., 1989, *Quantum Probability-Quantum Logic* (Lecture Notes in Physics, Volume 321), Berlin: Springer-Verlag.
- [Pit1994] Pitowsky, I. (1994). George Boole's 'conditions of possible experience' and the quantum puzzle. *The British Journal for the Philosophy of Science*, 45, 95-125
- [P1905] H. Poincare (1905), *Les mathematiques et la logique*. *Revue de metaphysique et de morale*, 815-835.
- [Po1938] Popper, K. R. (1938). A Set of Independent Axioms for Probability. *Mind*, 47(186), 275-277.
- [PS76] R. A. Purves and W. D. Sudderth, Some finitely additive probability, *Ann. Probab.* 4 (1976), 259-276.
- [RS2007] M. Rdei, S.J. Summers (2007): "Quantum probability theory" *Studies in the History and Philosophy of Modern Physics* 38, 390-417.
- [R1955] Rényi, A. (1955). On a New Axiomatic Theory of Probability. *Acta Mathematica Academiae Scientiarum Hungaricae*, 6:286-335.
- [S1976] Shafer, Glenn; *A Mathematical Theory of Evidence*, Princeton University Press, 1976.
- [S1974] Stengle, G. (1974). "A Nullstellensatz and a Positivstellensatz in Semialgebraic Geometry". *Mathematische Annalen* 207 (2), 87-97

- [ST1943] Shohat, J. A.; Tamarkin, J. D. (1943): "The Problem of Moments", New York: American mathematical society.
- [S97] Stinchcombe, M. B. Countably additive subjective probabilities. *Rev. Econom. Stud.* 64, 1 (1997), 125-146.
- [S16] Stinchcombe M. B. Objective and subjective foundations for multiple priors *Journal of Economic Theory*. 165: 263-291. (2016).
- [T44] Tarski, A. 1944, "The semantic conception of truth?", *Philosophy and Phenomenological Research*, 4 (3): 341-376.
- [TvD1988] Anne Sierp Troelstra, Dirk van Dalen, "Constructivism in Mathematics: An Introduction, Volume 1", 1988.
- [T2011] Tao, T., *Topics in random matrix theory*, American Mathematical Society, Providence, RI, 2011
- [T2012] Tao, T. (2012, April 2). A cheap version of nonstandard analysis [Blog post]. Retrieved from <https://terrytao.wordpress.com/2012/04/02/a-cheap-version-of-nonstandard-analysis/>
- [TZ2012] K. Tent, M. Ziegler (2012): *A Course in Model Theory*, Lecture Notes in Logic, vol. 40, Cambridge University Press, Cambridge.
- [V2012] J. Vinnen (2012): Second order logic or set theory? *Bull. Symbolic Logic* 18, no. 1, 91-121
- [V2016] S. Vineberg: *Dutch Book Arguments*, Stanford Encyclopedia of Philosophy, online publication: <http://plato.stanford.edu/entries/dutch-book/>
- [VP94] Plato, Jan von (1994): *Creating Modern Probability*. Cambridge: Cambridge University Press.
- [YH52] Yosida, K., and Hewitt, E. Finitely additive measures. *Trans. Amer. Math. Soc.* 72 (1952), 46-66.
- [W09] Alexander Wilce (2009): Test spaces. In K. Engesser, D. M. Gabbay, D. Lehmann, editors: *Handbook of quantum logic and quantum structures: quantum logic*, chapter 8, Elsevier, pp. 443-549

Appendix A. A contradictory set up

We present here a simple exercise taken from a widely used, application oriented, very high quality textbook. In the exercise, which is set up so as to mimic a realistic production problem, the authors propose a set of assumptions about probabilities and ask for the calculation of several other probabilities; the required calculations can be carried out without difficulties; it is very likely that this exercise has been solved thousands of times. What is problematic, however, is that on carrying out one extra calculation one realizes that the actual set of assumptions is inconsistent. In [AT2006], Exercise 2.8 page 67 presents the following problem.

Example 4. *On a given day, casting of concrete structural elements at a construction project depends on the availability of material. The required material may be produced at the job site or delivered from a premixed concrete supplier. However, it is not always certain that these sources of material will be available. Furthermore, whenever it rains at the site, casting cannot be performed. On a given day, define the following elements:*

E_1 = *there will be no rain*

E_2 = *production of concrete material at the job site is feasible*

E_3 = *supply of premixed concrete is available*

with the following respective probabilities: $P(E_1) = 0,8$, $P(E_2) = 0,7$, $P(E_3) = 0,95$ and $P(E_3|E_2^c) = 0,6$ whereas E_2 and E_3 are statistically independent of E_1 .

(a) *Identify the following events in terms of E_1, E_2 , and E_3 :*

(i) A = *casting of concrete elements can be performed on a given day;*

(ii) B = *casting of concrete elements cannot be performed on a given day.*

(b) *determine the probability of the event B .*

(c) *If production of concrete material at the job site is not feasible, what is the probability that casting of concrete elements can still be performed on a given day?*

We just briefly mention the intended solution. It is implicitly assumed, from the theory presented in the book, that all events are jointly perceivable.

(a) $A = E_1 \cap (E_2 \cup E_3)$ $B = A^c = E_1^c \cup (E_2^c \cap E_3^c)$;

(b) by the independence of (any combination of) E_2, E_3 from E_1 we have

$$\begin{aligned} P(B) &= 1 - P(A) = 1 - P(E_1 \cap (E_2 \cup E_3)) \\ &= 1 - P(E_1)P(E_2 \cup E_3). \end{aligned}$$

Since $E_2 \cup E_3 = E_2 \cup (E_2^c \cap E_3)$ and $P(E_2^c \cap E_3) = P(E_3|E_2^c)P(E_2^c) = 0,6 \times (1 - 0,7) = 0,18$, we have

$$P(B) = 1 - (0,8 \times (0,7 + 0,18)) = 1 - 0,704 = 0.296. \quad (16)$$

(c) we have

$$\begin{aligned} P(A|E_2^c) &= \frac{P(A \cap E_2^c)}{P(E_2^c)} \\ &= \frac{P(E_1 \cap (E_2 \cup E_3) \cap E_2^c)}{P(E_2^c)} \\ &= \frac{P(E_1 \cap E_3 \cap E_2^c)}{P(E_2^c)} \\ &= \frac{P(E_1)P(E_3 \cap E_2^c)}{P(E_2^c)} = \frac{0,8 \times 0,18}{0,3} = 0.48 \end{aligned}$$

Something, however, is not correct in this set up: we have $P(E_3 \cap E_2^c) = P(E_3|E_2^c)P(E_2^c) = 0.18$ so that

$$P(E_3 \cap E_2) = P(E_3) - P(E_3 \cap E_2^c) = 0.77 > 0.7 = P(E_2) \quad (17)$$

which contradicts monotonicity of P . Alternatively, again from (16), $P(E_3 \cup E_2) = 0.88 < 0.95 = P(E_3)$.

We conclude that the pre-environment described here is contradictory; the contradiction does not appear in the intended calculations, which were all obtained by sound applications of inference rules, but only with a careful choice of the events to examine.

As a consequence of the contradiction, for every statement, including the one in (16), the opposite can also be inferred: as $P(E_2 \cup E_3) \geq P(E_3) = 0.95$,

$$P(B) = 1 - (0,8 \times P(E_2 \cup E_3)) \leq 1 - 0,76 = 0.24 \neq 0.296. \quad (18)$$

Appendix B. Analysis of the contradictory setup

The set up described in Appendix A proposes a probability environment, which requires admissibility, a decidable question by Corollary 3.2. As an illustration, we determine once again that the requirements in Appendix A are contradictory directly by the algebraization method of Lemma 3.2 and Corollary 3.2.

First, observe that there are three events involved, so that $\mathbb{A} = \{E_1, E_2, E_3\}$. Then select a real variable for each of the boolean combinations on which conditions are given; it is convenient to use the following notation:

$$x_{\beta_1, \beta_2, \beta_3} = P(E_1^{\beta_1} \cap E_2^{\beta_2} \cap E_3^{\beta_3})$$

where $\beta_m \in \{-1, 0, 1\}$ and $A^{-1} = A^C$, $A^0 = \Omega$, $A^1 = A$. The equations expressing the requirements of the probability environment are the following, where we have assumed that the claimed independence is actually the full independence of the algebra generated by E_2 and E_3 from the algebra generated by E_1 (as it makes sense that the wheather is independent from any combination of human productions):

$$\left\{ \begin{array}{ll} x_{1,0,0} & = 0.8 \\ x_{0,1,0} & = 0.7 \\ x_{0,0,1} & = 0.95 \\ x_{1,1,1} & = x_{1,1,0} \cdot x_{0,0,1} \\ x_{1,-1,1} & = x_{1,-1,0} \cdot x_{0,0,1} \\ x_{-1,1,1} & = x_{-1,1,0} \cdot x_{0,0,1} \\ x_{-1,-1,1} & = x_{-1,-1,0} \cdot x_{0,0,1} \\ x_{0,-1,1} & = 0.6 \cdot x_{0,-1,0} \end{array} \right. \quad (19)$$

With the trivial substitution $x_{0,-1,0} = 1 - x_{0,1,0}$ there are 12 variables in the system.

Now make the change of variables

$$x_{\beta_1, \beta_2, \beta_3} = \sum_{\alpha_m \in \{\beta_m + |\beta_m| - 1, \beta_m - |\beta_m| + 1\}, m=1,2,3} y_{\alpha_1, \alpha_2, \alpha_3}$$

where $y_{\alpha_1, \alpha_2, \alpha_3}$ indicates the unknown probability of $E_1^{\alpha_1} \cap E_2^{\alpha_2} \cap E_3^{\alpha_3}$. After substitution and the inclusion of the conditions on the y_α 's the system has 9

equations of degree either 1 or 2, and 8 inequalities. As discussed here below there is no solution, indicating that the requirements are not admissible; a certificate that the system has no solution is in Appendix E.

One can actually wonder if it was the value 0.95 required for $P(E_3)$ which created a problem. In fact, one can leave $x_{0,0,1}$ as an indeterminate and solve the system in y_α for the other 7 variables. The result, before imposing the condition that $y_\alpha \geq 0$ for all α 's, is:

$$\begin{cases} y_{-1,1,1} &= (21 - 50y_{1,1,1})/50 \\ y_{1,-1,1} &= (12 - 25y_{1,1,1})/25 \\ y_{1,1,-1} &= 2y_{1,1,1}/3 \\ y_{-1,-1,1} &= (-3 + 10y_{1,1,1})/10 \\ y_{-1,1,-1} &= (21 - 50y_{1,1,1})/75 \\ y_{1,-1,-1} &= (24 - 50y_{1,1,1})/75 \\ y_{-1,-1,-1} &= (-3 + 10y_{1,1,1})/15. \end{cases}$$

With the nonnegativity condition one has $P(E_1 \cap E_2 \cap E_3) = y_{1,1,1} \in [3/10, 21/50]$, and $P(E_3) = y_{1,1,1} + y_{-1,1,1} + y_{1,-1,1} + y_{-1,-1,1} = 3/5$. This amounts to the cylindrical decomposition according to Tarsky-Seidenberg reduction theorem, although it is better computed by directly solving the equations first.

In conclusion, only the value $P(E_3) = 3/5$ is admissible in this set up.

Appendix C. Example

simple situation in which our theory applies.

Example 5. *Show that if five equiprobable collectively independent events are such that the probability of each exceeds that of the overall intersection by 0.5, then the probability that at least one event occurs can be bounded by the sum of the probabilities of any three of them (instead of all five as would follow from subadditivity). Show also that the probabilities of two of them are not enough in the case above, but they are if the excess is 0.55 instead of 0.5.*

Incorrect Solution C.1. *Let A_i , $i = 1, \dots, 5$ indicate the five events, and let $a = P(A_i) - P(\cap_{i=1}^5 A_i)$ be the indicated excess. We start from $a = 0.5$.*

Denoting $P(A_i) = z$, by inclusion-exclusion, independence and equiprobability, the probability that at least one event occurs satisfies

$$\begin{aligned}
p(z, a) &= P(\cup_{i=1}^5 A_i) \\
&= \sum P(A_i) - \sum_{i_1 \neq i_2} P(A_{i_1} \cap A_{i_2}) + \sum_{i_1 \neq i_2 \neq i_3} P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) \\
&\quad - \sum_{i_1 \neq i_2 \neq i_3 \neq i_4} P(A_{i_1} \cap A_{i_2} \cap A_{i_3} \cap A_{i_4}) + P(\cap_{i=1}^5 A_i) \quad (20) \\
&= 5P(A_1) - 10P(A_1)^2 + 10P(A_1)^3 - 5P(A_1)^4 + P(A_1)^5 \\
&= 6P(A_1) - 10P(A_1)^2 + 10P(A_1)^3 - 5P(A_1)^4 - a \\
&= 6z - 10z^2 + 10z^3 - 5z^4 - a.
\end{aligned}$$

1. We are then asked to prove that $p(z, 0.5) - 3z = 3z - 10z^2 + 10z^3 - 5z^4 - 0.5 \leq 0$ for $z \in [0, 1]$, which is easily shown by simple calculations. In fact, $p(z, 0.5) - 3z \leq 3z - 10z^2 + 10z^3 - 3z^4 - 0.5 = z(1-z)(3-7z+3z^2) - 0.5$, so it is sufficient to show that $q(z) = z(3-7z+3z^2) - 0.5 \leq 0$, but $q(0), q(1) < 0$, and $q(z)$ computed at the root of $q'(z) = 0$ is negative.
2. On the other hand, $p(0.3, 0.50) = 0.6295 > 2 \times 0.3$, so that $p(z, 0.5) \leq 2z$ does not hold for all $z \in [0, 1]$. Therefore, a bound by the probabilities of two events is not sufficient.
3. Finally, we see that once again $r(z) = p(z, 0.55) - 2z < 0$ for $z \in [0, 1]$; in fact, $r(0), r(1) < 0$, $r''(z) < 0$ as it is an irreducible second order polynomial, and if z^* indicates the only real root of $r'(z) = 0$, computable by solving a third degree polynomial, then $r(z^*) < 0$. So, if the excess a equals 0.55 then the probability of the union can be bounded by the sum of two of the probabilities of the single events.

Improved Solution C.2. It is possible to verify the existence of a probability environment (Ω, \mathbb{A}, P) satisfying the requirements by the algebraization method in Corollary 3.2. Clearly, for this simple case there is plenty of shortcuts, but, by way of exemplification, let's follow the abstract scheme.

There are no requirements on Ω and \mathbb{A} is required to be of size 5. So let $\mathbb{A} = \{A_i, i = 1, \dots, 5\}$, let $a = P(A_i) - P(\cap_{i=1}^5 A_i)$ be the indicated excess, and consider the boolean combinations $B_\beta = \cap_{i \in \beta} A_i$ for $\beta \subseteq \{1, 2, \dots, n\}$.

The stated requirements are then expressed by the following system

$$\begin{cases} x_{\{1,2,3,4,5\}} + a - x_{\{i\}} = 0, & i = 1, \dots, 5 \\ x_{\beta} - \prod_{i \in \beta} x_{\{i\}} = 0, & \beta \subseteq \{1, 2, \dots, n\} \\ x_{\{i\}} - x_{\{j\}} = 0, & i, j = 1, \dots, 5, i \neq j. \end{cases} \quad (21)$$

Perform the change of variables $x_{\beta} = \sum_{\alpha: \alpha_i=1 \text{ for } i \in \beta} y_{\alpha}$ and add the normalization and nonnegativity relations to get the system

$$\begin{cases} y_{\{1,1,1,1,1\}} + a - \sum_{\alpha: \alpha_i=1} y_{\alpha} = 0, & i = 1, \dots, 5 \\ \sum_{\alpha: \alpha_i=1 \text{ for } i \in \beta} y_{\alpha} - \prod_{i \in \beta} \sum_{\alpha: \alpha_i=1} y_{\alpha} = 0, & \beta \subseteq \{1, 2, \dots, n\} \\ \sum_{\alpha: \alpha_i=1} y_{\alpha} - \sum_{\alpha: \alpha_j=1} y_{\alpha} = 0, & i, j = 1, \dots, 5, i \neq j \\ \sum_{\alpha \in \{-1,1\}^5} y_{\alpha} - 1 = 0. \end{cases} \quad (22)$$

The nonemptiness of the semialgebraic set determined by the last system can be determined by cylindric reduction according to Tarsky-Seidenberg elimination, or shortened by some substitution. Indicating by z the common value of the $\sum_{\alpha: \alpha_i=1} y_{\alpha}$'s one gets to

$$\begin{cases} z^5 + a - z = 0 \\ \sum_{\alpha: \alpha_i=1 \text{ for } i \in \beta} y_{\alpha} - z^{|\beta|} = 0, & \beta \subseteq \{1, 2, \dots, n\} \\ \sum_{\alpha \in \{-1,1\}^n} y_{\alpha} - 1 = 0. \end{cases} \quad (23)$$

For $a = 0.5$ one can show by Sturm's theorem, or similar methods, that there are indeed solutions in $[0, 1]$ (see Remark C.3 below). Assuming z^* is one such solution, then one can take a concrete probability space with 5 independent events, each with probability z^* , which is known to exist; in such a space all requirements hold. Therefore, the probability environment is well defined and Part 1. of the incorrect solution is actually correct.

On the other hand, if $a = 0.55$ then by the same algebraic methods above, one can see that there is no solution of the first equation in $[0, 1]$ (see Remark C.3 below), and hence the assumptions about the excess being 0.55 are contradictory, and the entire calculation in Part 3. of the incorrect solution does not make any sense at all.

Notice that the equation $z^5 + a - z = 0$ cannot be solved by radicals for the given values of a ; hence, the result about the roots is purely existential.

Remark C.3. Notice also that Sturm's theorem gives a condition on a for the existence of a solution in $[0, 1]$ of $z^5 + a - z = 0$. The Sturm sequence in $z = 0$ is $a, -1, -a, 1 - \frac{3125a^4}{256}$ and in $z = 1$ is $a, 4, 4/5 - a, 1 - \frac{3125a^4}{256}$. Hence, there are two solutions in $[0, 1]$ for $a \in [0, \frac{4}{5^{5/4}})$, one for $a = \frac{4}{5^{5/4}}$ and none for $a > \frac{4}{5^{5/4}}$. Hence, the requirements are admissible if and only if $a \leq \frac{4}{5^{5/4}} \approx 0.535$.

There is yet another twist in Part 2. of the incorrect solution. There, we are trying to derive a negative result from the assumptions, so the consistency of the assumptions of the probability environment is not enough; a complete solution of the exercise uses the methods of Corollary 3.3

(Complete) Solution C.4. (to Example 5). According to the above corollary, we augment System (21) by the equation $x_{1\vee 2\vee 3\vee 4\vee 5} - 2x_{\{1\}} > 0$, where $x_{1\vee 2\vee 3\vee 4\vee 5}$ corresponds to $P(A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5)$. After the **x-y** change of variables, and the same substitutions as in the improved solution, we get System (23) incremented by the inequality $1 - (1 - z)^5 - 2z > 0$. Combined with $z - z^5 - 0.5 = 0$ this gives $(1 - z)^5/2 + z^5 < 0$, which is impossible to satisfy if $z \geq 0$. Since without the additional relation this is a probability environment, it follows from Corollary 3.3 that $P(A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5) \leq 2P(A_1)$ holds for $a = 0.5$, and that also Part 2. of the incorrect solution is wrong.

The point of the complete solution to Part 2. is that indeed $p(z, 0.5) \leq 2z$ does not hold for all $z \in [0, 1]$, but it holds for all the z 's for which the hypothesis make sense.

Appendix D. Example of Dutch Book

Example 6. An incorrect evaluator of the probabilities of two events A_1 and A_2 might assume that: they are independent, all their boolean combinations are jointly perceivable, $P(A_1|A_2) = 1/2$, and, finally, that $P(A_1) \neq 1/2$. Setting $x_{\beta_1, \beta_2} = P(A_1^{\beta_1} \cap A_2^{\beta_2})$ the above equirements give rise to a system with 4 terms

$$\begin{cases} x_{1,1} - x_{1,0}x_{0,1} & = 0 \\ x_{1,1} - \frac{1}{2}x_{0,1} & = 0 \\ x_{0,1} & \neq 0 \\ x_{1,0} - \frac{1}{2} & \neq 0. \end{cases}$$

After substitution with the y_α 's, one gets a system with 9 terms, $m_1 = 3$ equations, $m_2 = 4$ inequalities, and $m_3 = 2$ with \triangleleft replaced by \neq . A polynomial certifying that there is no solution is

$$\begin{aligned}
v &= \sum_{r=1}^2 t_r f_r + \prod_{r=8}^9 (h_r)^2 \\
&= \left((y_{1,1} + y_{-1,1})(y_{1,1} + y_{1,-1} - \frac{1}{2}) \right) \left(y_{1,1} - (y_{1,1} + y_{1,-1})(y_{1,1} + y_{-1,1}) \right) \\
&\quad + \left(-(y_{1,1} + y_{-1,1})(y_{1,1} + y_{1,-1} - \frac{1}{2}) \right) \left(y_{1,1} - \frac{1}{2}(y_{1,1} + y_{-1,1}) \right) \\
&\quad + (y_{1,1} + y_{-1,1})^2 (y_{1,1} + y_{1,-1} - \frac{1}{2})^2 \equiv 0.
\end{aligned}$$

Now following steps 1.-6. in the proof of the Theorem 4.1 for a fixed set of permutations, we get

$$\begin{aligned}
\tilde{t}_2 \tilde{f}_2 &= \left(-\mathbb{I}_{1,1,(\sigma_1^{(1,1)})} \mathbb{I}_{1,1,(\sigma_2^{(1,1)})} - \mathbb{I}_{1,1,(\sigma_1^{(1,1)})} \mathbb{I}_{1,-1,(\sigma_5^{(1,-1)})} + \frac{1}{2} \mathbb{I}_{1,1,(\sigma_1^{(1,1)})} \right. \\
&\quad \left. - \mathbb{I}_{1,1,(\sigma_1^{(1,1)})} \mathbb{I}_{-1,1,(\sigma_7^{(-1,1)})} - \mathbb{I}_{-1,1,(\sigma_7^{(-1,1)})} \mathbb{I}_{1,-1,(\sigma_5^{(1,-1)})} + \frac{1}{2} \mathbb{I}_{-1,1,(\sigma_7^{(-1,1)})} \right) \\
&\quad \cdot \left(\frac{1}{2} \mathbb{I}_{1,1,(\sigma_3^{(1,1)})} - \frac{1}{2} \mathbb{I}_{-1,1,(\sigma_8^{(-1,1)})} \right)
\end{aligned}$$

and \tilde{u} and V accordingly. One can see that there is no easy way of controlling the different copies which are used, except that of taking all permutations.

On the other hand, one can construct a simpler and more direct payoff by identifying the events as much as possible with the original A_i 's, rather than with their expansion in standard normal form, and by a clever choice of the copies. Here is a possible form (in which the number of the copy is indicated in the indices):

$$\begin{aligned}
V' &= \left(\mathbb{I}_{A_2,(1)} (\mathbb{I}_{A_1,(3)} - \frac{1}{2}) \right) \left(\mathbb{I}_{A_1 \cap A_2,(2)} - \mathbb{I}_{A_1,(4)} \mathbb{I}_{A_2,(2)} \right) \\
&\quad + \left(-\mathbb{I}_{A_2,(1)} (\mathbb{I}_{A_1,(3)} - \frac{1}{2}) \right) \left(\mathbb{I}_{A_1 \cap A_2,(2)} - \frac{1}{2} \mathbb{I}_{A_2,(2)} \right) \\
&\quad + \mathbb{I}_{A_2,(1)} (\mathbb{I}_{A_1,(3)} - \frac{1}{2}) (\mathbb{I}_{A_1,(4)} - \frac{1}{2}).
\end{aligned}$$

Since in each factor of each product there appear only events belonging to copies which are different from those appearing in the other factors (of the same product), independence can be used to show that $E(V') > 0$ for the incorrect evaluator; on the other hand, $V' \equiv 0$ as checked by simple algebraic expansion.

Appendix E. A short cut Dutch Book

Example 7. According to Corollary 4.3, it is possible to use a known contradiction to build a Dutch Book. We do it for the contradiction found at the end of Appendix A for the problem presented there, using the formalization in Appendix B. The chain of deductions leading to the contradiction in (17), in the notation of Appendix B is the following:

$$\begin{aligned}
x_{0,1,1} - (x_{0,0,1} - x_{0,-1,1}) &= 0 \\
x_{0,0,1} - 0.95 &= 0 \\
x_{0,-1,1} - 0.6x_{0,-1,0} &= 0 \\
x_{0,-1,0} - (1 - x_{0,1,0}) &= 0 \\
x_{0,1,0} - 0.7 &= 0 \\
x_{0,1,0} - x_{0,1,1} &\geq 0 \\
x_{0,1,0} - 0.7 &= 0
\end{aligned}$$

Notice that the fifth and the seventh equations coincide, as this relation appears twice in the reasoning expressing the contradiction. Now, we combine these relations with suitable coefficients in such a way that the single variables are telescopically canceled; we can use any real coefficient for all relations except the sixth, which needs a nonnegative coefficient. We indicate directly the random variables, as all relations are linear and there is no need of using copies; moreover, we directly replace the \mathbf{x} variables by a random variable, as we already expressed the properties of probability in equations number 1, 4 and 6. $\mathbb{I}_{i,j,k}$ is the random variable substituting the variable $x_{i,j,k}$. The payoff of the Dutch Book is

$$\begin{aligned}
\tilde{v} = & \frac{1}{0.07}[(\mathbb{I}_{0,1,1} - (\mathbb{I}_{0,0,1} - \mathbb{I}_{0,-1,1})) + (\mathbb{I}_{0,0,1} - 0.95) \\
& - (\mathbb{I}_{0,-1,1} - 0.6\mathbb{I}_{0,-1,0}) - 0.6(\mathbb{I}_{0,-1,0} - (1 - \mathbb{I}_{0,1,0})) \\
& + 0.6(\mathbb{I}_{0,1,0} - 0.7) + (\mathbb{I}_{0,1,0} - \mathbb{I}_{0,1,1}) - (\mathbb{I}_{0,1,0} - 0.7)] = -1.
\end{aligned}$$

Appendix F. Paradoxes

We make a brief reference to the so called paradoxes, many of which have been proposed concerning the foundations of probability theory (see, for instance, [E2012, Ha2013]). The point in most of the paradoxes in probability is that one either sets up contradictory collection of assumptions, or is forced by too rigid axiom systems to assign probabilities where there is no natural way of doing it. Both problems are addressed and solved by probability environments.

As just one example of how our formulation can deal with paradoxical settings, we consider Humphrey's paradox [Hu1985] and its variants [Ly2014], which turn out to be troubling for most foundations of probability. The paradox gives reasons to make assumptions about some events $B_{t_0}, I_{t_1}, T_{t_2}$, namely

- (i) $P(T_{t_2}|I_{t_1} \cap B_{t_0}) = p > 0$,
- (ii) $1 > P(I_{t_1}|B_{t_0}) = q > 0$,
- (iii) $P(T_{t_2}|\neg I_{t_1} \cap B_{t_0}) = 0$;

each variant gives then reasons to make an assumption about $P(I_{t_1}|T_{t_2} \cap B_{t_0})$, for instance that it equals $P(I_{t_1}|\neg T_{t_2} \cap B_{t_0}) = P(I_{t_1}|B_{t_0})$ in the first variant. It can be easily seen that the statements of the first three variants of the paradox are simply setting forth an inconsistent collection of requirements, easily detectable by algebraization. In each case, a Dutch Book can be produced against the believer of such assumptions: in the first variant, for instance, it can be obtained as follows. Let $\mathbb{I}_{A,j}, j = 1, 2$ be the indicator function of the event A in copy j , for two independent copies of the events. Then, by omitting the (mathematically irrelevant) indications of the times t_0, t_1, t_2 , let

$$\begin{aligned} V = & \mathbb{I}_{B,1}\mathbb{I}_{T,2}\mathbb{I}_{B,2}(1 - q) + (\mathbb{I}_{T,2}\mathbb{I}_{B,2}\mathbb{I}_{I,1}\mathbb{I}_{B,1} - \mathbb{I}_{B,1}\mathbb{I}_{T,2}\mathbb{I}_{I,2}\mathbb{I}_{B,2}) \\ & + (\mathbb{I}_{T,2}\mathbb{I}_{B,2}q\mathbb{I}_{B,1} - \mathbb{I}_{T,2}\mathbb{I}_{B,2}\mathbb{I}_{I,1}\mathbb{I}_{B,1}) - \mathbb{I}_{B,1}\mathbb{I}_{T,2}(1 - \mathbb{I}_{I,2})\mathbb{I}_{B,2}; \end{aligned}$$

we have

$$\begin{aligned}
E(V) &= P(B)P(T \cap B)(1 - q) \\
&\quad + (P(T \cap B)P(I \cap B) - P(B)P(T \cap I \cap B)) \\
&\quad + (P(T \cap B)qP(B) - P(T \cap B)P(I \cap B)) - P(B)P(T \cap I^c \cap B) \\
&= P(B)P(T \cap B)(1 - q) \\
&\quad + (P(T \cap B)P(I \cap B) - P(B)P(T \cap B)) \\
&\quad + (P(T \cap B)qP(B) - P(T \cap B)P(I \cap B)).
\end{aligned}$$

From the requirements we know $P(T \cap B) \geq P(T \cap I \cap B) = pP(I \cap B) > 0$ and $P(B) > 0$; the first term is strictly positive, and all the other terms in the r.h.s. of the last equation are 0; hence from the requirements, $E(V) > 0$. On the other hand, a simple expansion shows that $V \equiv 0$. Therefore, V is a weak Dutch Book, witnessing the inconsistency of the requirements.

The fourth variant of the paradox suggests to leave $P(I_{t_1}|T_{t_2} \cap B_{t_0})$ as "undefined". We can translate this assumption into the setting of probability environments by stating that no requirement is made on such probability. In such case, the other requirements are consistent, and the paradoxical nature of the example disappears; a probability environment exists, and the last probability can then be computed as necessary consequence of the other requirements in such environment, necessarily taking the value $P(I_{t_1}|T_{t_2} \cap B_{t_0}) = 1$ (as easily seen by algebra or by Bayes formula).

All of the most relevant formalizations of Probability Theory seem to be affected by Humphrey's paradox, as they force us to assign a value to $P(I_{t_1}|T_{t_2} \cap B_{t_0})$ before hand, hence entering in one of the three contradictory variants. Only Rényi's axiom system allows to leave such probability as undefined, and hence is able to "respond" to the paradox, reaching the same conclusion as with probability environments.

In a further extension of the paradox (see [Ly2014] Section 6.6), though, one can set up things in such a way that also Rényi's axiom system would be forced to assign values which should remain undefined; the issue is immaterial for probability environments, which then solve extension of the paradox as well.

Aknowledgments

The author would like to thank P. Dai Pra, M. Maggesi, R. W. J. Meester, D. Mundici and C.C.A. Sastri for valuable discussions and comments.